Article

UC EA University Council for Educational Administration

Educational Administration Quarterly
1–42

DOI: 10.1177/0013161X16687006
journals.sagepub.com/home/eaq
SAGE

Administrator Observation and Feedback: Does It Lead Toward Improvement in Inquiry-Oriented Math Instruction?

Jessica G. Rigby[1], Adrian Larbi-Cherif[2], Brooks A. Rosenquist[2], Charlotte J. Sharpe[2], Paul Cobb[2], and Thomas Smith[3]

Abstract

Purpose: This study examines the content and efficacy of instructional leaders' expectations and feedback (press) in relation to the improvement of middle school mathematics teachers' instruction in the context of coherent systems of supports. **Research Method/Approach:** This mixed methods study is a part of a larger, 8-year longitudinal study in four large urban school districts across the United States. We used transcripts of interview data, surveys, and video recordings of instruction of 271 cases, over 4 years, to determine the content of administrator press, as reported by teachers, and the relationship between the content and change (if any) in instruction. To do so we used qualitative coding of interview transcripts, and ran a series of statistical models to examine the nature of the variance in and impact of administrative press. **Findings:** Most of the administrators'

[1]University of Washington, Seattle, WA, USA
[2]Vanderbilt University, Nashville, TN, USA
[3]University of California, Riverside, CA, USA

Corresponding Author:
Jessica G. Rigby, University of Washington, College of Education, Box 353600, Seattle, WA 98195-3600, USA.
Email: jrigby@uw.edu

press, as reported by teachers, was not targeted toward specific teachers' mathematics instruction in ways that would likely lead toward improvement in those practices. Rather, the press focused on content-neutral instructional practices or classroom management and organization. **Implications for Research and Practice:** The instructional leadership practice of administrator observation and feedback is widespread, yet understudied as it relates to changes in teacher practice. Our findings indicate that current policies that mandate principals to spend substantial time in classrooms are unlikely to result in significant improvements in the quality of instruction unless meaningful resources are invested to support administrator learning.

Keywords
school administration, observation and feedback, mathematics instruction, instructional leadership, mixed methods, empirical paper

Introduction

School administrators' actions are important for what happens in classrooms (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; Coelli & Green, 2012; Hallinger & Heck, 1998; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Robinson, Lloyd, & Rowe, 2008; Supovitz, Sirinides, & May, 2010; Waters, Marzano, & McNulty, 2003). Historically, for administrators to engage in the work of instructional improvement was to cut directly against institutionalized organizational arrangements that long buffered teachers from leaders' feedback (Meyer & Rowan, 1977; Murphy, Hallinger, & Heck, 2013). However, in recent decades, the role of school-site administrators[1] has shifted from organizational manager to instructional leader, responsible for the support of teacher learning (e.g., Bryk et al., 2010; Louis, Leithwood, Wahlstrom, & Anderson, 2010). While there is no one agreed-upon definition of what it means to be an instructional leader (Rigby, 2014), there is general consensus that administrators' instructional leadership responsibilities include attending to the teaching and learning in classrooms (e.g., Council of Chief State School Officers, 2015).

The change of administrators' roles from manager to instructional leader has been accompanied by concurrent organizational shifts in states, central offices, and schools to increase high-quality learning opportunities for students. As early as the late 1980s and early 1990s, state policy makers in California worked to provide consistent instructional policy messages by aligning curriculum guidance, text adoption, and assessments (Cohen & Hill, 2008). At the district level, extensive studies of both New York City District #2 and San Diego Unified School District in the late 1990s and early 2000s

found that a change in central office focus from compliance to instructional issues that included coherent supports in the form of professional development for teachers, centralized curriculum, and culture change resulted in increased student achievement (Darling-Hammond et al., 2003; Elmore & Burney, 1997; O'Day & Quick, 2009). Similarly, Newmann, Smith, Allensworth, and Bryk (2001) studied Chicago Public Schools' elementary schools and found that schools that had greater instructional program coherence, or a set of interrelated programs guided by a common framework and sustained over time, had greater student achievement gains. More recently, Cobb and Jackson (2011a) called for a coherent system of supports at the district level, specifically oriented toward ambitious mathematics instruction that includes job-embedded professional development along with coach support, school leaders' practices as instructional leaders in mathematics, and district leaders' practices to support capacity building for instructional improvement at the school level. Finally, Bryk et al. (2010) put forth a framework of supports at the school level that describes how professional capacity, school learning climate, parent, school, and community ties, and instructional guidance all interact with classroom instruction. At any of these three organizational levels, it is unlikely that a single support on its own will lead toward instructional improvement at scale, rather organizational systems must be redesigned (e.g., Elmore, 1996; Hopkins & Woulfin, 2015).

This mixed methods study took place in four large urban school districts that were seeking to improve the quality of middle school mathematics instruction and student learning opportunities through coherent systems of supports. As part of the districts' plans for instructional improvement, school administrators were expected to be instructional leaders who both communicated expectations to teachers for quality instruction and gave math-specific feedback to math teachers. In this analysis, we seek to first, understand the nature of press that teachers reported receiving using qualitative analyses of interview transcripts; second, learn the extent to which enacted administrative press is associated with improvement in the quality of inquiry-oriented math instruction through quantitative analyses; and third, consider the types of support that may enable administrators to improve their instructional leadership practices. We consider these questions in the context of coherent systems of support, although in this analysis we narrow our focus on one of the supports: the role of the school leaders. In what follows, we review the applicable current literature on middle school mathematics and leadership practices, outline the context of this study, put forth a conceptual frame that considers the variety of both instructional and leadership practices, review our mixed methods, describe our findings, and finally discuss the significance and implications of the findings.

Middle School Mathematics

We are particularly interested in the context of mathematics, in part due to the gulf between best practices and actual instructional practice. Inquiry-oriented mathematics instruction, or that which promotes student investigation and development of conceptual understanding, is not commonplace in middle school classrooms (see, e.g., Hiebert & Stigler, 2000). Rather, the more normative approach of instruction involves teaching mathematical procedures easily replicated on standardized tests that most often require a procedural-level of thinking (Stein, Grover, & Henningsen, 1996; Thompson, Philipp, Thompson, & Boyd, 1994).[2]

Research in mathematics education has long pointed to a need for more rigorous student learning goals in math. The widespread uptake of these goals has come as the result of research in the learning sciences over the past few decades, which demonstrates the promise of inquiry-oriented instructional practices (Hiebert, 1997; Kazemi, Franke, & Lampert, 2009; Stein, Engle, Smith, & Hughes, 2008). Such practices include selecting cognitively demanding math tasks (Stein et al., 1996), launching the task to support students in engaging with the mathematical relationships in the task (Jackson, Garrison, Wilson, Gibbons, & Shahan, 2013), supporting groups of students as they work toward constructing a solution, and orchestrating concluding whole-class discussions in which teachers scaffold presentations in order of mathematical sophistication (Stein et al., 2008). These practices, the body of which have been called "ambitious" math instruction by Lampert and others (e.g., Lampert, Beasley, Ghousseini, Kazemi, & Franke, 2010), are distinct from more prevalent traditional practices in that they support students to engage in cognitively demanding tasks and construct solutions to tasks that require students to make sense of the mathematical relationships at hand. Teaching ambitious mathematics, then, requires a deep understanding of the mathematical content, an understanding of the various ways in which students solve tasks for a given topic, an understanding of what students currently know and can do, an understanding of common errors that students make for the designated topic, and the ability to anticipate and sequence solution strategies for whole-class discussions (Ball, Thames, & Phelps, 2008; Stein et al., 2008).

Despite the push toward ambitious instruction, many studies of U.S. math instruction show that teachers typically select tasks and materials that ask students to reproduce mathematical procedures presented by the teacher (Hiebert et al., 2003; Schoenfeld, 2002; Stein et al., 1996). Even in districts where teachers are expected to use curricula with cognitively demanding tasks, the implementation and the curriculum are often not aligned (Cohen,

1990; Remillard, 1999). Instead, teachers often demonstrate procedures for solving the tasks and then give students time to practice these skills, instructional practices similar to a traditional classroom and curriculum. Teachers' development of instructional practices involves significant learning on their behalf, a reorganization of practice rather than a mere extension or elaboration of current practice (Ball & Cohen, 1999; Gamoran et al., 2003; Kazemi et al., 2009). In short, teachers must reimagine and relearn what it means to teach mathematics. Significant school- and district-level supports are essential to engender such teacher learning and reorganization of practice. Among a coordinated system of supports, school leader press may serve as an ongoing lever for instructional improvement. For administrator feedback to be effective, it is likely just as important that administrators reimagine and relearn what it means to teach mathematics as it is for teachers.

Observation and Feedback for Instructional Improvement

One long-standing common instructional leadership routine is classroom observation and feedback (Horng, Klasik, & Loeb, 2010; May & Supovitz, 2011; Supovitz et al., 2010). An organizational routine is a repetitive, recognizable pattern of interdependent actions, carried out by multiple actors (Feldman & Pentland, 2003). In this case, the routine consists of an administrator observing classroom instruction, taking notes, and providing either verbal or written feedback to the teacher. There are a wide variety of approaches to taking notes on instruction, including scripting (Halverson, Kelley, & Kimball, 2004), using "look fors" (Protheroe, 2009), and more recently tablet-based forms both available from companies and developed by individual school districts.

Although the observation and feedback routine is widely implemented, it is understudied. Specifically, there is little research that has sought to understand the extent to which administrator observation and feedback is a mechanism for instructional improvement. Despite the dearth of research, the routine's widespread adoption is grounded in a robust conceptual argument for why it may support teachers' improvement of their instructional practices. First, there is a growing body of evidence that indicates that teachers' engagement with colleagues who are more accomplished in activities directly relevant to their instructional practices can support development of ambitious instruction (Coburn & Russell, 2008; Elmore, 1996; Frank, Zhao, & Borman, 2004; Louis, Marks, & Kruse, 1996; Newmann, King, & Youngs, 2000; Penuel, Riel, Krause, & Frank, 2009). Second, there is evidence that teachers' opportunities to learn are amplified when new insights are actionable and relate directly to their classroom practice (Ball & Cohen, 1999). Blase and Blase (1999) argue

that principals can use informal, ongoing feedback following observations to promote reflection, communicate goals for classroom instruction, and encourage teacher growth by promoting professional development and teacher collaboration. The authors argue that these activities may lead to improved instruction. Third, school leaders are also teachers' evaluators. While the confluence of seemingly conflicting roles is not without tension (Darling-Hammond, Wise, & Pease, 1983; Glanz, 2005; Ovando & Ramirez, 2007; Zepeda, 2007), the added evaluative press might provide impetus for teachers to improve their practice. Finally, school leaders' communication of appropriate instructional expectations can be important in the context of a system that also includes a variety of supports for instructional improvement, including instructional coaches and professional development (Cobb & Jackson, 2012). In this case, teachers might receive aligned support from a number of sources: feedback from the principal, support from an instructional coach, and professional development from the central district office.

In addition to the conceptual arguments for the potential of the observation and feedback routine to promote instructional improvement, there is also some research in the field of educational leadership as a part of larger studies of instructional leadership broadly defined. The research is largely quantitative, and aims at understanding the impact of principal practice on student achievement. For the most part, the assessments these studies use to measure student learning emphasize procedural skills and are not aligned with ambitious math instruction (Cobb & Jackson, 2011b). A smaller group of studies addresses the connection between principals' actions and change in teacher practice, as measured by either student test scores or instructional improvement. The studies that explored the impact of school leadership on student achievement had few significant findings. For example, Shin and Slater (2010) found that there was not a relationship between student achievement and either the time that principals spent on instructional leadership (defined as developing curriculum and pedagogy) or the time spent on supervising and evaluating teachers; Horng et al. (2010) found no relationship between multiple school outcomes (including student achievement) and time devoted to instructional tasks, such as classroom observation; and a follow-up study found that time spent on informal classroom observations, or walkthroughs, was negatively associated with learning and school improvement (Grissom, Loeb, & Master, 2013). While these studies suggest that there is little reason to believe that school leader observation and feedback might lead toward improvement in student outcomes, they investigated instructional leadership writ large by assessing relationships between inputs and outputs rather than the mechanisms through which the observation/feedback routine might influence teachers' daily practices.

In contrast to the research in educational leadership, there is extensive research on the efficacy of feedback on performance in the fields of educational psychology and organizational behavior that suggest that the observation/feedback routine may in fact improve teacher practice. For the most part, this research focuses around three elements of feedback: who gives it, the content of the feedback itself, and when it is given (Brinko, 1993; Scheeler, Ruhl, & McAfee, 2004; Van Houten, 1980). Of these elements, the literature points to the content as the most salient. In particular, teachers were more likely to change their practice if they received more specific feedback. Shute (2008) argues that this may be because of the reduction in uncertainty that may decrease cognitive load, and lead to both more motivation and more efficient strategies to complete a task (p. 157). In his research on teachers' feedback to students, Van Houten (1980) echoes this sentiment, "The more precise the feedback, the more rapid the learning" (p. 52). Hilberg, Waxman, and Tharp (2004) also argue that the content of feedback from classroom observations is essential to instructional improvement. They contend that teachers are unaware of what they need to change in their practice. Feedback, then, helps "teachers understand their own strengths and weaknesses and has consequently enabled them to improve their instruction significantly" (p. 9). The authors cite several studies that found changes in teacher practice when they were given "appropriate" feedback and suggestions regarding their individual instruction. Finally, Scheeler et al. (2004) found that more specific feedback led to an increase in targeted instructional behaviors. There were not conclusive findings on either who gives the feedback (e.g., the principal or a peer) or the timing of feedback, although Scheeler et al. (2004) found in their review of 10 studies that immediate feedback led to faster learning. It is evident that teachers need expert support and that school-based administrators are currently situated as a main lever of influence on teachers' instructional practice.

While the above research points to school administrators' potential to provide meaningful feedback to teachers toward instructional improvement, there are several institutionalized challenges to realizing this potential, especially in middle schools. First, there is a historical and institutionalized norm that separates classroom instruction from school administration (Lortie, 2009; Meyer & Rowan, 1977; Wolcott, 1973). Although the accountability movement and No Child Left Behind have opened the doors of classrooms to administrators, the decoupling of instruction from administrative leadership might persist. Second, in the accountability era, some administrators are likely to press for instructional practices that will lead to proximal success on standardized test scores rather than for instructional practices that support students' development of conceptual understanding (Spillane et al., 2002). Third, research on teachers' learning indicates that coparticipation in activities that are close to

practice with a colleague who has already developed relatively accomplished instructional practices is essential if teachers are to develop inquiry-oriented instructional practices (Ball & Cohen, 1999; Borko, Jacobs, Koellner, & Swackhamer, 2015; Grossman, Hammerness, & McDonald, 2009; Sztajn, Borko, & Smith, in press). Thus, observation and feedback might not be sufficient to support teachers' reorganization of their current practices even when the feedback is of high quality. Adding to the challenge, administrators in the middle school level may have only limited familiarity with the discipline-specific content that is the focus of instruction (Louis, Dretzke, & Wahlstrom, 2010). Finally, as mentioned above, there is ample research that highlights the tension between the dual roles of administrators as both evaluators and instructional supports (Darling-Hammond et al., 1983; Glanz, 2005; Ovando & Ramirez, 2007; Zepeda, 2007). This tension may prevent teachers from being able to view feedback as support for instructional change, even if the feedback is of high quality.

Our study addresses the gaps in what we know about the observation/feedback routine, in the context of the urgent need to improve teachers' instructional practices. Specifically, we investigated the content of teacher reports of administrator expectations and feedback, or *press*, to see if variations in types of expectations and feedback correlated with teachers' instructional practices as measured by the Instructional Quality Assessment (IQA, described in more depth below; Boston, 2012), or if there was a "match" between the aspects of instruction that teachers' needed to improve and the press administrators gave. In the context of districts that aimed to support teachers' development of inquiry-oriented instruction through a coherent system of supports, we asked the following research questions:

Research Question 1: To what extent do middle school teachers report administrator feedback that focuses on inquiry-oriented mathematics instruction?
Research Question 2: Do administrators vary the content of their feedback based on teachers' mathematics instruction?
Research Question 3: Is there a relationship between administrator feedback and improvement in teacher instruction?

Our findings suggest that even in districts who aim for coherent systems of support, administrators' expectations and feedback, as described by teachers, were not targeted toward specific teachers' mathematics instruction in ways that would likely orient improvement in those practices. In other words, for the most part, there was not a "match" between the administrator press and teacher needs. Across 4 years in four school districts, we found that the

content of only a few administrators' press was likely to communicate appropriate expectations for how teachers might improve their practice; and administrator press was not associated with changes in teacher practice. These findings have both practical and theoretical implications. First, they have the potential to guide future work in the field of education leadership as we rethink the supports administrators might need in order to engage in this routine in ways that could lead to substantial improvement in instruction in mathematics. Second, the findings suggest that the administrators' lack of pedagogical content knowledge in mathematics hindered their potential to substantially improve teachers' instructional practices.

Although the data in this study predate the implementation of the Common Core State Standards in Mathematics (CCSS-M), the standards are relevant to this study. The goals for students' mathematical learning detailed in the CCSS-M are consistent with the inquiry-oriented approach described above (Cobb & Jackson, 2011b; Hiebert & Grouws, 2007). Many states have either adopted the CCSS-M standards or have developed standards influenced by them that emphasize both conceptual understanding and procedural fluency as important learning goals. The findings from this study, set in districts that had goals for mathematics instruction aligned with the CCSS-M, are, therefore, likely relevant to the challenges currently faced by school leaders across the country.

Conceptual Frame

In any given school building, teachers are likely to have a range of teaching experience and skill. So while school leaders may have similar expectations for all their mathematics teachers, it is expected that in order to improve their practice, individual teachers need feedback specific to their current instruction. For example, a first-year teacher may need feedback on classroom management and organization more than a veteran teacher who may need help learning how to connect student responses during a discussion of students' solutions. In other words, productive administrative press will be matched with teachers' instructional needs. We operationalize this through both elements: types of teachers' instruction, and types of administrators' press.

Types of Teachers' Instruction

We categorized the possible array of teachers' instruction along a scale we call Instructional Bands. This scale is aligned with research on ambitious mathematics (e.g., Kazemi et al., 2009; Stein et al., 2008) but it is at the level of detail that is relevant to school administrators. The four Instructional

Bands indicate the next steps in teachers' learning, and the types of press that would be appropriate to support their learning. Each band suggests the need for a different type of press for improvement. Descriptions of the bands are as follows:

- *Traditional*: The teacher does not select conceptually rigorous tasks but instead uses tasks that emphasize the application of a procedure to produce a correct answer without making connections to deeper mathematical concepts or meaning.
- *Proceduralized*: The teacher selects cognitively demanding tasks with multiple solution methods, but then transforms the tasks into a procedural activity, thereby lowering the cognitive demand and restricting the number of potential solutions.
- *Low-level discussion*: The teacher selects cognitively demanding math tasks and allows the students to explore the mathematical concepts. However, the teacher does not lead a concluding whole class discussion in which students are pressed to explain their reasoning and connect their solutions to different solution methods. Students might describe their solutions but do not justify why their solutions are valid.
- *Ambitious*: The teacher selects cognitively demanding tasks, allows the students to explore the mathematical concepts, and leads a whole class discussion in which students are pressed to explain their reasoning and connect different solutions. Although some phases of the lesson may not be exemplary, there are opportunities for rigorous student learning throughout.

Types of Administrators' Press

While some general school leader press may be useful across all four bands (such as questioning techniques), relying on the research on productive feedback, we argue that to significantly support teacher development of inquiry-oriented mathematics instruction, teachers need specific press directed at their mathematics instruction, or "matched" press. We conceptualize this match as feedback that is likely to support improvements in teacher instruction in ways that are discernable as measured by the Instructional Bands described above. We categorize types of press on two dimensions: the content of mathematical press, and the content of nonmathematical press.

The content of mathematical press would vary across the four Instructional Bands. For example, a teacher with traditional practices may receive feedback to press him to choose a high-level task, whereas a teacher with ambitious practices may receive feedback that presses her to attend to making

connections between multiple representations of a concept. This type of press is likely to lead to instructional improvement toward ambitious practices.

We anticipated that we would find at least two other categories of nonmath press: *classroom management and organization*, or press that is not directly about the content of instruction; *general instruction*, or press that is focused on instruction, but is not discipline specific. First, while teachers must manage and organize their classrooms, these characteristics are not sufficient to significantly reorganize mathematics instruction. So while teachers' instruction may improve with administrator press about classroom management, it is unlikely that this type of feedback will support teachers' development of inquiry-oriented instruction. Second, administrator press for general instruction may be useful to improve instruction in any discipline, including mathematics. This type of press may support teachers with most types of instruction, as general instructional foci may improve inquiry-oriented mathematics instruction. For example, a key practice in inquiry-oriented mathematics instruction is students working together to solve a task (Stein & Lane, 1996). An administrator may press a teacher to include more "group work" in their lessons, so that students can engage with each other around mathematical ideas. We posit that this type of press is appropriate for all teachers except those who have developed ambitious practices, who will likely only improve by receiving press that is specifically about mathematics.

Figure 1 illustrates how we conceptualize the match between press and teacher practice categorized by Instructional Band.

Note that there is a check mark (✓) in the General Instruction row rather than a star (★). This is to indicate that while there is the possibility of this type of press improving teachers' practice, it is more likely that press focused specifically on mathematics will significantly reorganize teachers' practice in the ways needed to meet the learning demands of ambitious mathematics. Recognizing that ambitious instruction and feedback for improvement are both complex activities, this conceptualization is necessarily simplified (due to the nature of our data and the study as a whole). However, these broad strokes are useful to gain a beginning understanding of the routine, an addition to the current field of educational leadership.

It is important to note that this study takes place in potentially productive contexts, or school districts in which school leaders were expected to be instructional leaders in mathematics as a part of a broader system of supports. This is an essential element of our conceptual frame; it is likely that appropriate/matched school leader press will contribute to teachers' development of ambitious instruction within the context of a system of coordinated supports for teachers' learning (Cobb, Jackson, Smith, Sorum, & Henrick, 2014). This implies that both teachers and administrators receive

	Classroom Management & Organization	General Instruction	Mathematics
Traditional		✓	★
Proceduralized		✓	★
Low-level Discussion		✓	★
Ambitious			★

Figure 1. Match between press and instructional band.
✓ = press has potential to improve teacher inquiry-oriented instructional practice
★ = press has the most potential

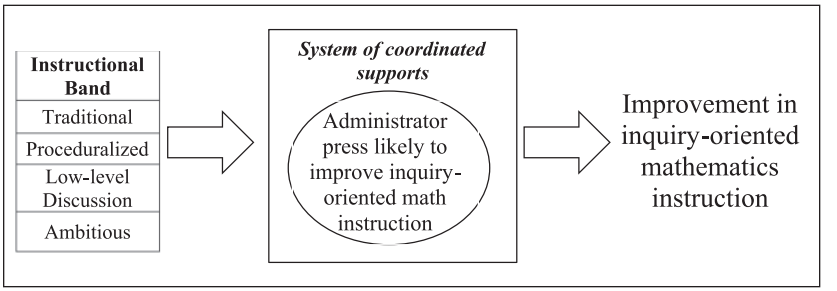

Figure 2. Administrator press is likely to contribute to teachers' development of inquiry-oriented math instruction.

professional development on inquiry-oriented mathematics, district leaders have an explicit vision for inquiry-oriented mathematics instruction, and teachers have access to high-quality inquiry-oriented curriculum and receive ongoing support in implementing it effectively. Our conceptual frame is illustrated in Figure 2.

Data Sources and Method

This research is a part of a larger study that investigates what it takes to improve middle school mathematics instruction at the scale of a large urban district in the United States, Middle School Mathematics and the Institutional Setting of Teaching (MIST). From 2007 to 2011, multiple forms of data were collected from four large urban school districts to test and refine hypotheses and conjectures about how new positions, learning events, organizational routines, and tools might support teachers to improve instruction (Cobb & Jackson, 2012; Cobb & Smith, 2008). The research is situated in specific settings—four school districts that were chosen because they

appeared to have more developed, coordinated systems of supports than one would expect in a randomly sampled U.S. district. First, all four districts espoused a specific theory of action around mathematics instruction: they aimed to support teachers' development of ambitious instructional practices, and three of the four adopted a curriculum consistent with this goal (the fourth district created their own inquiry-oriented curriculum). Second, the districts both expected administrators to communicate expectations for math instruction in addition to observe instruction and provide feedback to leverage instructional improvement, and they also provided administrators with professional development to learn how to do so. In each of the four districts, principals (and sometimes assistant principals) attended district-sponsored professional development sessions on both high-quality mathematics instruction (sometimes with their teachers or math coaches), as well as on observation and feedback routines. While the research is scant, there is evidence that school leaders can learn content-specific leadership practices through professional development (Barth, 1986; Boston, Henrick, Gibbons, Berebitsky, & Colby, 2016; Carver, Steele, & Herbel-Eisenmann, 2010; Nelson, 1997; Sassi & Nelson, 1999). Given these districts' ambitious goals for instructional improvement and infrastructures of support (discussed in detail below), they presented promising sites for exploring how administrators might work as part of a coherent instructional system to support teachers' development of ambitious practice (Cobb & Jackson, 2011a). Thus, understanding how school administrators in these contexts worked with mathematics teachers illuminates how leaders may serve as one lever in a system of coordinated supports toward instructional improvement.

During the years that the data for this study were collected, the school districts in our sample implemented a range of organizational support structures aimed at ambitious goals for instructional improvement. Three of the school districts (A, B, and D) implemented Connected Mathematics Projects II (CMP2), an ambitious inquiry-oriented curriculum, and the fourth district (C) developed its own inquiry-oriented instructional materials. We classify both curricula as "ambitious," that is, they (a) provide cognitively demanding tasks (Stein et al., 1996); (b) allow students to explore mathematics from multiple entry points and with multiple solution paths; and (c) conclude investigations with a whole class discussion during which the teacher presses students to make connections between solution strategies and the underlying mathematics. The selection and implementation of additional supports varied across the four districts (see Table 1), but included ongoing professional development, regularly scheduled time during the school day for mathematics teachers to collaborate, and collaboration with mathematics instructional coaches.

Table 1. District Means[a] for Teachers' Survey Response of Instructional Supports Received.

	District A	District B	District C	District D	
	(1)	(2)	(3)	(4)	
	Mean (*SE*)	Mean (*SE*)	Mean (*SE*)	Mean (*SE*)	ANOVA Test of Equal Means, *p* Value
Hours PD	24.89 (1.53)	24.24 (1.40)	22.48 (1.33)	21.47 (1.92)	.394
Hours weekly in teacher collaboration	0.65 (0.17)	1.69 (0.29)	3.34 (0.18)	1.24 (0.39)	<.001
Frequency of principal observation	2.79 (0.45)	7.01 (0.99)	3.58 (0.41)	3.94 (0.55)	<.001
Frequency of principal feedback	1.16 (0.16)	5.29 (0.85)	2.44 (0.29)	3.18 (0.49)	<.001
Extent of coach assistance (0-4)	0.98 (0.15)	2.14 (0.12)	1.52 (0.22)	2.03 (0.14)	<.001
Probability school has a coach	0.46 (0.07)	1.00 (.)	0.58 (0.07)	0.87 (0.04)	<.001
N (teachers)	25	28	28	24	

[a]Values were first averaged within teacher, across years, and then teacher-specific averages were averaged within district.

Table 1 illustrates substantial differences in supports received by teachers across districts, such as teachers in District C engaged in relatively greater number of hours in teacher collaboration, and teachers in District B reported more frequent observation by and feedback from their principal. Teachers in Districts B and D were most likely to have access to and use assistance from instructional coaches. Due to these variations and our research focus on the potential influence of school administrators' influence on instruction, we controlled for district membership in our statistical analyses.[3]

While this analysis focuses on administrator press as a lever to improve instruction, the other elements of the larger system of supports are also important to consider. Table 1 highlights the various supports and their relative engagement across districts, and other analyses by MIST researchers examine the efficacy of these levers, such as coaches, teacher workgroups, and professional development.[4] All the currently completed analyses of the various system supports found gaps between the districts' theories of action toward improvement and implementation. For example, five separate analyses of the quality of teacher workgroups found that teacher learning opportunities were in some way limited, due to the type of data used in conversations, principal press for improvement on test scores rather than conversations about ambitious mathematics practices, and a lack of facilitative expertise (Horn et al., 2017; Horn, Garner, Kane, & Brasel, 2016; Horn, Kane, &

Table 2. Policies Delineating Administrator Responsibilities as Instructional Leaders Over Math.

Description of Policy	District A	District B	District C	District D
Communicate instructional expectations aligned with high-quality mathematics instruction	✓	✓	✓	✓
Informally observe mathematics teachers and provide feedback that supports instructional improvement	✓	✓	✓	✓
Provide time in the school schedule for mathematics teachers to meet	✓	✓	✓	✓
Attend teacher collaborative time	✓	✓		
Collaborate with school or district math coaches or department heads		✓	✓	✓
Either principal or an assistant principal can evaluate and support math department		✓		✓
Use data to assess effectiveness of instruction		✓		

Wilson, 2015; Kane, 2017; Rigby, Andrews-Larson, & Chen, 2014). A study examining quality and efficacy of professional development found that district facilitators initially struggled to provide learning opportunities for teachers to engage with conceptual ideas, instead focusing on the forms of the reform. Over 4 years, with expert support, the facilitators improved their adult learning practices and led professional development sessions in ways that allowed teachers to engage deeply with the conceptual ideas in teaching and learning ambitious mathematics (Boston et al., 2016). Administrators as a lever of support, then, are situated in the imperfect and often messy process of instructional improvement.

In all four districts and throughout the study, school administrators were expected to press and support teachers in learning how to enact inquiry-oriented mathematics instruction. The four central offices had these explicit expectations of their middle school administrators: (a) communicate expectations to teachers aligned with inquiry-oriented mathematics instruction, (b) observe mathematics instruction and provide feedback that supported instructional improvement, and (c) provide time in the school schedule for math teachers to meet. Across all four districts, school administrators were expected to spend at least 2 hours a day in classrooms observing and giving feedback. While we did not specifically collect these data, principal reports on district expectations ranged from the 2-hour minimum to "80% of my time in classrooms" (Principal B5, personal communication, January 27, 2010). Table 2 compares the districts' expectations for administrators as instructional leaders over mathematics.

As each district developed their own school administrator professional development agendas, they were distinct from one another. Three of the four districts worked with the University of Pittsburgh's Institute for Learning (http://ifl.pitt.edu/). This organization provided professional development in observation and feedback through their "Learning Walk" routine that addressed the central characteristics of an ambitious lesson from a discipline-neutral standpoint, such as supporting classroom discourse and academic rigor across content areas. In addition, all four of the districts provided some amount of specific professional development on mathematics, including what an ambitious mathematics lesson includes, and what to "look for" when conducting observations and providing feedback. The amount of this ongoing math-specific support varied, however, from a 3-hour session during a 2-day summer principal session to monthly sessions during principal meetings. Notably, District C had the least consistent professional development focused on mathematics, whereas District D had the most, including 3-day–long sessions in the third year of the study provided by MIST. Table 3 illustrates the range of school administrator professional development across the four districts.

The data for this analysis come from the first 4 years of the larger study that focuses on what it takes to improve middle school mathematics instruction at scale in large urban school districts in the United States. In these years, we followed approximately 120 teachers and 60 of their administrators per year; these educators were nested in 30 schools and across four school districts. Table 4 includes the descriptive statistics for our teacher sample.

In this analysis, we draw primarily on transcripts from teacher interviews and video recordings of classroom instruction that were previously scored using rubrics from the IQA (Boston, 2012). We used the IQA data to assign teachers to particular Instructional Bands (described below). The IQA was developed to provide both researchers and district personnel information about the quality of instructional practice at a single point in time. The instrument operationalizes inquiry-oriented instruction and thus also indicates areas in which teachers can improve instruction. The IQA is an appropriate tool for assessing instruction in this study given the focal districts' goals for mathematics instruction and student learning, and their use of inquiry-oriented mathematics curriculum (see Table 5).

Additionally, we used data collected from an annual survey given to teachers and school leaders every year in March. Among other topics, the survey asked questions about frequency of administrator observations, as well as descriptive information such as years of experience, courses taught, and so on. As we examined the change in teachers' practice in relationship to administrators' press over time, we limited our sample in this analysis to teachers

Table 3. Description of Administrator Professional Development, by District.

Description of Professional Development	District A	District B	District C	District D
Training by University of Pittsburgh's Institute for Learning	✓	✓	✓	
Summer sessions focused on math	✓	✓		✓
Ongoing PD through the year	✓(3-5 days)	✓(Occasionally)		✓(Monthly)
Other	✓(Some principals went to summer training offered by CMP)		✓(PD varied across regions, some embedded in other topics like RTI)	✓(In Year 3, MIST provided three sessions on supporting ambitious math instruction)

Table 4. Descriptive Statistics of the Sample.

Demographic Descriptor	*N* (105 total)	Percentage
Female	69	65.71
Grade_6[a]	41	39.05
Grade_7[a]	56	53.33
Grade_8[a]	58	55.24
Experience teaching math (years)		
1-2	28	26.67
3-5	18	17.14
6-10	18	17.14
11-15	17	16.19
16-20	13	12.38
21-30	7	6.67
>30	4	3.81
Race/ethnicity		
African American or Black	30	28.57
Hispanic	9	8.57
White (non-Hispanic)	63	60.00
Other	3	2.86

[a]Some teachers taught multiple grade levels; therefore, the sum of teachers teaching in sixth, seventh, or eighth grade is greater than the total sample size (*N* = 105).

Table 5. Data Analyzed in This Study.

Construct/Variable	Operationalization	Data Source
Lesson quality	Instructional band	Annual video recordings of two of lessons per participating teacher, coded with the Instructional Quality Assessment rubrics (Boston, 2012)
Administrator expectations	Teacher reports of administrator's expectations for mathematics instruction	Annual teacher interview
Administrator feedback	Teacher reports of administrator's feedback from classroom observation(s)	Annual teacher interview
Frequency of feedback from administrators following classroom observation	Teacher-level reports of frequency of administrator feedback following classroom observation	Annual teacher survey
Teacher experience	Teacher self-reports of years of experience	Annual teacher surveys

for whom we have at least two consecutive years of IQA data, and whom the same school administrator supervised for both years. Each year of data is considered a case; we have at least two, and up to four, cases per teacher. This resulted in approximately 70 teachers per district, with a total sample size of 271 cases over the first 4 years of the study.

Teacher cases excluded from our analysis were due to attrition: either that of the teacher or of the school administrator. For example, if a teacher participated in our study for three consecutive years, but a change of school leadership occurred in the third year, we would only include first 2 years of data in our analysis. Similarly, if a teacher participated in one year of the study, but not for a second (because they left our study, the school, the district, or the profession), we did not include their case. A comparison of our sample and all the excluded cases show that teachers generally reported receiving very similar levels of supports, such as coaching, frequency of administrator observation and feedback, and time for teacher collaboration. However, when teachers left our study, they were generally replaced with teachers with less experience. Additionally, teachers with less experience were more likely to leave their position (and thus, leave our study). Because teachers with less experience tended to enter and leave our study at higher rates, and because our sampling methods required consecutive years of participation, teachers retained in our analytical sample tended to have slightly more experience, on average, than teachers excluded from our analysis. However, years of experience did not affect overall IQA scores, there were no statistically significant differences between our analytic sample and the study sample as a whole. Below, we describe both the IQA instrument and the interviews.

Instructional Quality Assessment and Instructional Bands

As described earlier, the IQA is a measure of mathematics instructional quality in relation to inquiry-oriented instruction. The instrument was developed at the Learning Research and Development Center at the University of Pittsburgh (Boston, 2012). The IQA is based on Institute for Learning's Principles of Learning and assesses the cognitive demand and conceptual complexity of instruction along multiple dimensions. We determined that the IQA would be an appropriate tool for our purposes because it categorized the kinds of instructional practices that participating districts attempted to develop. Results of a pilot study conducted in urban middle schools indicate that the reliability of IQA for assessing the quality of mathematics lessons is good ($\alpha = .89$; percent of interrater agreement = 82% overall; Matsumura, Garnier, Slater, & Boston, 2008).

We video-recorded participating teachers' instruction over two consecutive days in January and February of each year. While teachers were expected to teach the content they would normally teach, we asked that they include a problem-solving activity and a related whole-class discussion in their instruction. These elements were compatible with the lesson structures in all four of the districts' curricula, as well as matching the districts' articulated goals of instructional reform. As a part of the larger MIST study, these videos were coded by a set of trained and reliable coders (percent agreement = 70.5%; for details on the coding, see, Jackson, Garrison, Wilson, Gibbons, & Shahan, 2013) using a set of seven rubrics (for links to the rubrics and more detailed information about all MIST instruments, see http://peabody.vanderbilt.edu/departments/tl/teaching_and_learning_research/mist/mist_instruments.php). For this analysis, we used the scores from the previously scored videos.[5]

The Instructional Bands are an ordinal scale developed to purposefully simplify the complex array of 12+ scores generated through IQA coding process in a way that preserves the vision and goals of inquiry-oriented instruction. The bands represent categories of instructional quality characterized by the cognitive demand of a teacher's task, maintenance of that cognitive demand when implementing the task, and the level of cognitive demand represented in the students' and teacher's discourse in the lesson's concluding discussion. The bands were constructed by aligning the coding schemes of three related IQA rubrics (Academic Rigor of the Task, Academic Rigor of Task Implementation, and Academic Rigor of the Discussion) and developing categories of teachers' practices at a grain-size that school leaders are likely to attend to during the observation and feedback routine.

Teacher Interviews

To understand teachers' perception of the content of administrator press, we analyzed transcripts of annual teacher interviews that were conducted each January. Administrator press is only a small section of a longer interview protocol, so we focused our coding on the portion of the interviews in which we asked about how they work with their administrators.[6] The questions focused on principal expectations for math instruction, principal support to meet these expectations, and details about the observation and feedback routine, including the content of the feedback received. This study operates under the assumption that the teachers' description of their administrator press is an adequate representation of the press they receive overall in the study year. We also looked at interviews conducted with principals in January of each year to understand their reports of their instructional expectations and the feedback they gave teachers. However, for our final analysis we decided

that what teachers considered salient and remembered were the elements that were most likely to have affected their instructional practices rather than how the administrator may have described his or her press. For that reason, we do not include the analyses of the principal interviews.

In this analysis, we are unable to distinguish between informal observation and feedback routines from formal evaluation processes. Teachers reported on feedback they received from their school administrator, and most often did not delineate which type of observation the feedback came from. We argue this discernment is not essential for our analyses for two reasons. Currently, there is a focus on teacher evaluation as a mechanism to improve teacher practice, as well as to evaluate teachers. This is seen in the frameworks themselves, such as the CEL 5D and the Danielson Framework for Teaching, the professional development provided to implement them, as well as research done on the efficacy of teacher evaluation on student outcomes (Taylor & Tyler, 2012). Second, in survey results from our sample, 63% of teachers reported that they viewed their administrator's observations to "evaluate my teaching" (n = 172) and 75% reported that they viewed the purpose to "assist me in improving my teaching" (n = 204). This provides further evidence of the strong overlap of the perceived purpose of the observation and feedback routine.

Data Analysis

Our analysis had four distinct phases, three of which were linked to our research questions and an additional fourth phase to explore possible alternate explanations.

Phase I. To develop measures of the content of administrator press, we analyzed transcripts of annual teacher interviews, as described above. We approached the initial wave of data analysis by reading 10% of the data corpus with an unbounded framework of a priori codes meant to both (a) help us begin to categorize types of feedback according to theory and research on teacher learning and instructional improvement and (b) allow for the development of more clear and consistent coding rules as we collected more examples. We then developed inductive codes from the data. The combination of using both deductive codes from extant literature and inductive codes that arose from our initial coding led to a final codebook that was attentive to the contexts and viewpoints in which our participants work while also attuned to current research and literature.

Very few teachers described feedback or expectations that were specific to their work as *math* teachers, and even fewer gave examples of feedback that

was specific to their mathematics instruction in specific lessons. For this reason, we developed the following coding scheme for administrator press (Table 6).

Our coding unit was by teacher response. As long as the teacher discussed the same topic, with or without interviewer probing, we considered responses and follow-ups to be a part of the same coding unit. Each coding unit could be coded multiple times, with the exception of a response coded *indistinct*. If we identified any other code, indistinct was not used. Most individual transcripts had multiple coding units, and therefore had several codes.[7] The data were coded using NVivo 10 by three coders, and a fourth coder (the first author) double-coded 20% of the data to ensure reliability and to serve as an "anchor coder" in instances of uncertainty flagged by the other coders. Reliability was maintained through ongoing conversations and double-coding.[8] When instances arose which necessitated further clarification or revisions to the codebook, all coders were notified and reviewed their coding to ensure that the data were all coded alike. This process iterated for the duration of the sample, or 271 transcripts. After all the data were coded, all discrepancies settled, and all flagged items resolved by the anchor coder, the data were exported from NVivo as quantitative output to be used in STATA for the quantitative analysis described in Phase II.

Phase II. To answer Research Question 1, "To what extent do middle school teachers report administrator feedback that focuses on inquiry-oriented mathematics instruction?" we examined the administrator press data in two ways. First, we calculated the frequency of mentions of different types of press and the percentage of teachers who mentioned each type (Table 7). This focuses on the "mention" level and provides us with an understanding of the variation in forms of press that administrators provided. Second, we calculated the frequency of the distribution of press, where teachers who reported multiple forms of press in a particular year were recoded according to type of administrator press. This measure suggests the number of teachers that reported the kinds of content-specific feedback that could help improve the rigor of their lessons.

To answer Research Question 2, "Do administrators vary the content of their feedback based on teachers' mathematics instruction?" we ran two sets of analyses. First, we used descriptive statistics to examine the relationships between the type of press teachers reported in a particular year and their practice as categorized by Instructional Band. We did a chi-square test for homogeneity on these results to examine if administrators varied their press based on teachers' instruction.

Table 6. Codes and Examples of Administrator Press.

	Code	Definition	Examples from the Data
Nonmath press	Indistinct	*Press did not describe focus on aspects of instruction.*	"He wants me to teach math," "she wants me to increase student test scores," "he wants the students to improve," and "he wants us to collaborate with each other." Also included in this category were instances when the teacher used a "buzz word" or name of a program (IFL, Accountable Talk, Rigor and Relevance) without describing what the term meant In addition, we used this code if the teacher said the administrator did not give feedback.
	Classroom management/ organization	*Press has potential to be useful for improving classroom management and organization.*	"Put your objectives on the wall," "students are/should be engaged," "have a word wall," noting the percentage of students who were on task, accurately following the pacing guide, increasing student test scores by a certain percentage, and developing relationships with students and parents.
	General instruction	*Press focuses on the teacher's instructional practices from a content-neutral stance.*	[He wants to see] "Students talk[ing] to each other," "students work[ing] together in groups," "using data to know what to reteach," making sure that students are engaged meaningfully in classwork, using formative assessment, and using questioning strategies to get students to think critically. One teacher reported that his principal expects " . . . collaboration with the students, he loves to see group work, he loves to see hands-on type activities . . ."
Math press	General mathematics	*Press is about math instruction, but general to all math teachers rather than specific to the teacher's instruction.* Note that there is a relatively low bar for teacher reports to be coded in this category, a simple statement relating to math was coded General Mathematics.	Examples include expecting for teachers to teach the inquiry-oriented curriculum (CMP), using inquiry-based instruction, expecting the introduction of the lesson to prepare students for the mathematics content of the lesson, asking students to justify their answers to math problem solutions, and including a real-world application of specific math concepts.
	Lesson-specific math feedback	*Press is particular to the individual teacher's instruction and specific to mathematics.* Note that this category has a relatively high bar for inclusion, teachers need to describe their particular lessons and the specific feedback the administrator gave in relation to their instruction.	Examples include principals giving an example of how they approached a particular lesson when they taught math, a principal pressing on a particular approach a teacher took for a whole class discussion, a principal commenting that a teacher lowered her expectations for a class by teaching "traditional" math rather than inquiry-oriented, and a principal noticing that students had a misconception and telling the teacher to reteach the concept.

Note. IFL = University of Pittsburgh's Institute for Learning; CMP = Connected Mathematics Project.

Second, to examine whether teachers' practices in different Instructional Bands were more or less likely to receive math-focused press, we conducted a series of hierarchical logistic regression models. For these analyses, we created an outcome binary variable called *math press* that was coded 0 if teachers reported only indistinct, classroom management/organization, or general instruction press. We coded *math press* as 1 if teachers reported any press related to general mathematics or lesson-specific mathematics feedback. As the observations of press and instruction were observed over multiple years for each teacher, and teachers nested within schools, the hierarchical models allow us to account for the clustering of observations within teachers and teachers within schools. In the models, district membership is treated as a fixed effect. The sample selection criteria (teacher in same school with the same administrator for multiple, consecutive years) excluded the possibility of teachers having multiple administrators within the same school. Thus, it was not necessary to run cross-classified models.

Instructional Band was coded as a series of dummy variables, with the traditional category suppressed (i.e., the coefficients associated with proceduralized, low-level discussion, and ambitious are relative to traditional).[9] To control for between-district differences that might be associated with both press and instruction, district dummy variables were included in the model. District C was selected as the suppressed category as it had the most infrequent reports of mathematics press. We ran a likelihood ratio test to examine whether instruction, as measured by Instructional Band, explained whether teachers were more or less likely to receive any math press.

Phase III. To answer Research Question 3, "Is there a relationship between administrator feedback and improvement in teacher instruction?" we ran a three-level hierarchical ordinal logistic regression model to examine if math press predicted change in Instructional Bands between 2 years. To do so, we regressed current Instructional Band level (and ordinal variable from traditional, proceduralized, low-level discussion, to ambitious) on math press and nonmath press while holding constant prior Instructional Band level (dummy coded with traditional the suppressed category). As in the previous model, district fixed effects are included to control for unobserved differences in district contexts that might be correlated with both instructional change and press.

Phase IV. Finally, we conducted an exploratory analysis using principal and teacher interview data for four administrators, whose teachers reported that they gave specific math press, to try and explain why these administrators were able to give specific math press. These data included principal interview

Table 7. Percentage of Administrator Press by Type.

Type of Press		*N* (teachers)	Percentage of Teachers
Nonmath press	Indistinct	194	71.6
	Classroom management/ organization	223	82.3
	General instruction	144	53.1
Math press	Math-specific	56	20.7
	Lesson-specific math	5	1.8
	Total	271	

transcripts and school-wide teacher reports on the principals' press in the specific study year.

Findings

Research Question 1: To what extent do middle school teachers report administrator feedback that focuses on inquiry-oriented mathematics instruction?

The largest percentage of teachers, 82%, reported that their administrators' press addressed classroom management and organization. In contrast, only 1.8% of teachers described administrator press that addressed an issue of mathematics particular to their classrooms. In total, just less than one quarter of the sample reported receiving any math press, with 20.7% reporting General Math Press and 1.8% reporting Lesson-Specific Math Feedback. Table 7 reports the number and percentages of types of press (note that teacher reports of press fall into multiple categories if they included more than one type of press, and each transcript had a number of coding units, often with distinct codes).

As Table 7 illustrates, the vast majority of administrator press was not specific to mathematics (77.5% of all reported instances of press). While administrator press on classroom management/organization and general instructional practices may improve instruction, we argue that in order to support teachers' development of sophisticated ambitious practices, teachers also need press and support that is directly related to the teaching of mathematics. Our coding of instances of administrator press suggests that, for the most part, administrators tended to give feedback focused on relatively easy-to-measure/observe aspects of classroom instruction (e.g., the presence of a

Table 8. Percentage Distribution of Type of Administrator Press by Instructional Band and Match.

	Administrator Press	Traditional	Proceduralized	Low-Level Discussion	Ambitious
Nonmath press	Indistinct	3.7%	7.61%	4.35%	7.69%
	CM/CO	37.04%	31.52%	15.22%	32.69%
	General instruction	40.74%[a]	42.39%	45.65%	36.54%
Math press	Math	16.05%[b]	16.30%[b]	32.61%	23.08%
	Lesson-specific math	2.47%[c]	2.17%[c]	2.17%[c]	0.00%[c]
	Total	81	92	46	52

[a]Press has potential to improve teacher inquiry-oriented instructional practice. [b]Press has more potential. [c]Press has most potential.

word wall, student knowledge of the learning target) rather than attending to the improvements in instruction that the teachers needed to support deeper student conceptual understanding. Furthermore, on the whole feedback was not specific to the content of instruction, which likely reduced the effectiveness of the observation and feedback routine to support teachers in developing ambitious instructional practices.

Research Question 2: Do administrators vary the content of their feedback based on teachers' mathematics instruction?

Table 8 shows the percentage distribution of type of administrator press for each category within the Instructional Bands. Overall, teachers' practices that were in the two most inquiry-focused Instructional Bands reported receiving more math press than the teachers whose practices were characterized in the less sophisticated two Instructional Bands.

These findings are consistent with what we find in the next analysis. To control for between-district contexts that may confound the bivariate relationships reported above, we ran a series of hierarchical logistic regressions models that included district fixed effects (see Table 9).[10]

Results from Model 1 (district fixed effects) show large differences in math press across districts. The odds of a teacher receiving math press were more than four times greater in District D as compared with District C. The likelihood of reporting math press in Districts A and B were similar, both

Table 9. Hierarchical Logistic Regression of Math Press on Instruction.

	District Fixed Effects		Instructional Band	
Fixed Effect	*OR*	*SE*	*OR*	*SE*
District A	2.44*	1.32	2.12	1.21
District B	2.59*	1.36	2.53*	1.33
District D	4.78***	2.50	4.63***	2.43
Instructional Band				
Traditional			1.12	0.48
Low-level discussion			2.30*	1.05
Ambitious			1.05	0.50
_cons	0.10	0.05	0.09	0.04
Random Effects	Variance	*SE*	Variance	*SE*
School level	1.32×10^{-34}	2.73×10^{-18}	2.60×10^{-34}	5.50×10^{-17}
Teacher level	0.49	0.56	0.36	0.52

Note. *OR* = odds ratio; *SE* = standard error. Coefficients reported as *OR*s.
*$p < .1$. **$p < .05$. ***$p < .01$.

more than twice the odds of District C. In the second model, we added the Instructional Band categories. A likelihood ratio test comparing Model 1 with Model 2 was statistically significant ($p < .05$), suggesting that Instructional Band level is associated with the likelihood of receiving math press. The coefficient on low-level discussion (odds ratio [*OR*] = 2.30, standard error [*SE*] = 2.05, $p < .1$) indicates that in our sample, teachers in the low-level discussion band (rigorous task that is maintained through the lesson, but a nonrigorous discussion) have twice the odds of receiving math press as teachers coded in the proceduralized band. This difference was statistically significant at the .10 level ($p = .06$). However, teachers with ambitious practices were no more likely to report press than teachers in the proceduralized band when controlling for district membership.

Research Question 3: Is there a relationship between administrator feedback and improvement in teacher instruction?

We conducted a hierarchical ordinal regression to examine the likelihood of instructional improvement (moving up the Instructional Band in the next year if the teacher reported receiving math press from his or her administrator). Results are presented in Table 10.[11]

Table 10. Ordinal Regression Modeling Change in Instructional Band.

Instructional Band	Odds Ratio	SE
Prior band		
Traditional	0.94	0.73
Low-level discussion	1.57	0.72
Ambitious	3.31**	1.62
Math press	1.24	0.45
District A	10.74***	6.94
District B	3.78***	1.89
District D	5.42***	3.02
Random Effects	Variance	SE
School	6.49×10^{-34}	2.49×10^{-17}
Teacher	0.15	0.60

Note. *SE* = Standard error.
*$p < .1$. **$p < .05$. ***$p < .01$.

Table 10 shows that math press was not related to being in a higher band in the current year when controlling for prior instruction and district membership. Not surprisingly, teachers who had ambitious practices in the prior year had three times the odds of maintaining ambitious practices than a teacher who had proceduralized instructional practices in the prior year. Teachers who reported math press were more likely to have lessons in the higher bands when controlling for prior instruction and district membership, but this result is not statistically significant. As a robustness check, we also conducted a fixed effects analysis at the teacher level (i.e., modeling change from no math press to math press over the course of year and movement up the Instructional Band in the following year) and did not find a statistically significant coefficient for math press. This finding is consistent with the results in Table 10 and provides further evidence that teachers who reported math press were not more likely move to a higher band in the subsequent year.

Exploratory Analysis: Principal Characteristics

Finally, we investigated whether characteristics of the administrators influenced the type of administrator press. Our principal sample size is too small to run statistical analyses ($n = 29$), but descriptive statistics show that 90% (26) of the principals gave math press over the time they were in the sample. Of those, 14% (4) gave lesson-specific math feedback. The remaining three

principals (10%) who did not give math press gave general instruction press. As we were unable to detect much variation using quantitative methods, we explored the qualitative data for further explanation.

We focused on the five cases of administrators whose teachers who reported giving lesson-specific math feedback (a total of four principals, one teacher reported receiving this type of feedback in both Years 3 and 4). To do so, for each case we read a summary of the principals' interview (which includes their descriptions of what a high-quality math lesson looks like, descriptions of what they look for when they observe classrooms, and what type of feedback they give to teachers) and a summary of all the interviews in that school that year (which synthesizes the interviews with all the teachers, the administrators, and the coach, if applicable).

Overall, the main similarity among these principals was their district and exposure to the math curriculum: three of the four were in District B (the fourth was in District A); and all but one teacher report occurred in either the third or fourth year of the study. With the one exception (the report from the first year of the study), the timing indicates that the principals had exposure to 3 or 4 years of CMP2 and the associated professional development, as well as the press from the district central offices toward ambitious math instruction. However, this was true in all four of the districts and for many other principals who had a similar exposure but did not give as sophisticated feedback. Otherwise, there were no similarities common to all four principals in our data. Two principals had math experience; one had a background as a high school history teacher, and the other PE. Three principals were in their 3rd year at their school, one was in his 12th year.

Discussion

Research on teacher learning in math education has demonstrated that teachers need ongoing support to significantly improve their practice. The districts in our sample conceptualized administrative press as one lever for improving instruction among a coordinated system of supports, and provided administrator professional development on both conducting this routine, albeit not always math specific, and on ambitious mathematics instruction. However, our findings indicate that most of the administrators in our sample did not differentiate the kind of feedback they gave in ways that would support teachers as learners. The majority of the administrative press was not specific to inquiry-oriented math instruction. Rather, the majority of the teachers in our sample reported receiving press mostly on easily observable aspects of instruction and classroom management, such as the presence of a learning objective or standard posted on the board, or the number of students engaged

in the classroom. While these aspects of instruction are important, these kinds of feedback are not, by themselves, likely to support teachers' development of inquiry-oriented practices. The predominance of content-neutral feedback suggests that the districts' push and professional development around ambitious math instruction only marginally affected this aspect of the administrators' instructional leadership.

One interesting finding was that teachers whose practice had low-level discussions had a higher likelihood of receiving math press at a statistically significant level ($p > .1$), whereas teachers with ambitious practices did not have this higher likelihood. This finding could indicate that school administrators did not feel compelled to give math press to teachers with ambitious practice because of their sophisticated instruction; they may have not had the expertise to know how to improve already ambitious instruction.

These findings deserve serious consideration for several reasons. First, they can be taken as an indicator of the capacity of school leaders, against which the field's expectations for leadership practices seem extremely ambitious. To be instructional agents of instructional improvement in mathematics and perhaps also in other content areas, administrators need to improve the quality of press. Our findings indicate that this is a challenging task: It was not the norm even in districts with atypical, reform-oriented agendas in math instruction and ambitious expectations for and professional development to support administrators as instructional leaders. If we continue to put our trust in observation and feedback as a central instructional leadership support, districts need to consider the learning needs of school administrators rather than seeing the routine as a method of enforcing compliance.

What forms of support would such an overhaul entail? We identify two possible paths and, based on our findings, recommend one. The first path is to build administrator content expertise. Looking to teacher education and professional development literature for effective models of support, we conjecture that administrators need opportunities to first, receive professional development that is math-specific and organized around the math instructional materials that teachers are using, and second, use that knowledge to engage in pedagogies of investigation and pedagogies of enactment (Grossman, Compton, et al., 2009). That is, they should have opportunities to consider problems of practice (specifically, the targeted forms of math instruction and their role as leaders in supporting it) through reflection on their own and modeled practice, discourse, and guided rehearsals of their own (Horn, 2010). We outline this path given that the observation and feedback routine is a deeply rooted responsibility of school leaders. If the routine is to continue, it needs to be enacted with greater expertise.

The second and recommended path is to assign this work to available math experts (and, indeed, other content experts as well) while the role of the administrator as instructional leader is redefined. Rather than attempting to improve administrators' capacity in multiple content areas, school districts could take a systems-based improvement approach, identify and leverage existing capacity, while simultaneously identifying and supporting individual learning needs. For example, a principal could focus on teachers who struggle with management and basic instructional skills, whereas a math coach could work with the teachers who are attempting to enact more sophisticated mathematics-specific instructional strategies.

Recent research on the role of the administrator as an instructional leader points away from practices such as the observation and feedback routine, and toward coordinated support for teacher learning within schools (e.g., Bryk et al., 2010; Horng & Loeb, 2010; Louis, Leithwood, et al., 2010). In this conception of the principal as the instructional leader, the routine of observation and feedback would be the responsibility of a school's entire instructional leadership team. In this scenario, the observation and feedback routine could be used by the principal to collect information to collaboratively drive formal learning events, such as teacher workgroups or professional development, rather than relying on the routine to be a main lever to improve instruction. The literature on teacher learning indicates that teachers are more likely to develop ambitious practice if the "learning events" (Cobb & Jackson, 2012) are organized around a small number of high-leverage practices, and guided by a more expert practitioner (Grossman, Hammerness, & McDonald, 2009). In this case, administrators could gather information across classrooms to use in collaboration with an instructional team to design learning events. Second, administrators too may benefit from working with a more expert other, such as the mathematics coach, to use the observation/feedback routine as a high-leverage practice in which they themselves learn from coaches about how to support and direct teachers' work.

This recommendation reflects the view that instructional leadership at a school comprises a series of functions that may be carried out by a variety of forms, such as members of the leadership team. That is, the principal should not need be the only instructional leader in the school. Rather, we propose that instructional leadership practices, approaches, and expertise that may be coordinated by, but not carried out entirely by the school leader. Principals already share their leadership in a number of ways, such as by dividing teacher evaluations among assistant principals (if available), working with coaches to help struggling teachers, and by giving teacher leaders formal facilitative roles in Professional Learning Communities. Yet principals are currently trained, supported, and evaluated as if they must perform all the

leadership functions in a school as an individual (National Policy Board for Educational Administration, 2015). The findings of this study suggest that the current formal and informal policies around school site administration set school leaders up for failure. They cannot do it all and, as a consequence, enact the forms of the leadership actions but not their functions (e.g., Spillane & Callahan, 2000). We acknowledge that this is not a "costless" alternative. Coaches and other leaders also need to build their expertise and content knowledge, as well as skill in giving appropriate feedback. However, given school administrators' current workload that is both broad and deep, this distributed approach may lead to higher quality learning opportunities for more students.

Limitations

While the qualitative data provided a more detailed description of the type of press teachers reported receiving from administrators, the data itself has limitations. First, teachers were instructed to include a problem-solving lesson and a whole-class discussion for the video-recorded lessons (which were later scored using the IQA rubrics). While these elements were a part of both their curriculum and the districts' articulated visions for reform, it is possible that the lessons were teachers' "best shot" at enacting the instructional practices rather than their typical practices. It could therefore be the case that the teachers' typical lessons would be scored lower on the IQA than their recorded scores. However, the IQA scores are already relatively low, and when taken in context with the rest of the study's data (including a test for Math Content Knowledge and audio recordings of teacher workgroups), we are confident that the video-recorded lessons were reasonably representative of the teachers' typical instruction.

Second, our interest in the content of administrator press necessitated a reliance on teacher interview data, which introduces the problem of recollection and self-report. Teachers were interviewed in January, but asked about the feedback they had received over the school year. A second limitation of these data was in the changing nature of the interview both by design and by circumstance. By circumstance, and as consequence of working on a large research team, different interviewers often probed to different degrees because of their own differences in background or occasionally in response to circumstances interrupting or otherwise abbreviating the interview. To account for these differences, we maintained a "low ceiling" for what counted in each of our analytical codes, with the exception of lesson-specific math press.

Self-reported data inherently involves risks in terms of both validity and reliability. We addressed these challenges in several ways. First, to address

issues of validity, we used interview data rather than survey data. In the interviews, we were able to ask follow-up probes for clarification that was not possible in the survey data. As discussed above, we also maintained a low ceiling as what counted in each of our analytical codes. This also addresses the effect of potential interviewer variability on validity. To address issues of reliability, we triangulated our data across teachers who received feedback from the same administrator as well as across years. We typically saw similar reports from teachers across years and within schools (Tourangeau, Rips, & Rasinski, 2000).

Finally, as described in our methods, our sampling technique was more likely to exclude newer teachers. As newer teachers might benefit the most from administrators' press, we may have captured a bias sample in terms of the lack of instructional improvement in relationship to administrator press. We accounted for this possibility by examining our sample to the study sample as a whole and found that years of experience did not affect overall IQA scores, and there were no statistically significant differences between our analytic sample and the study sample as a whole.

Implications

Based on research that firmly places school leaders as key lever for instructional change (Bryk et al., 2010; Coelli & Green, 2012; Hallinger & Heck, 1998; Leithwood et al., 2004; Robinson et al., 2008; Supovitz et al., 2010; Waters, Marzano, & McNulty, 2003), we argue that their actions as instructional leaders are necessary, but not sufficient, to support teachers' improvement in ambitious teaching practices. Rather, they must be a part of coherent systems of supports. Given the long history of the challenge of improving instruction at scale (Elmore, 1996), it is not entirely surprising that neither the administrators nor the other supports met the high expectations of the districts' theories of action. However, the findings provide key insight to the nature of and implications for one key element in instructional reform at scale, the role of instructional leaders, and therefore, provide meaningful implications for both practice and policy.

For Practice

Our data suggest that the participating administrators spent much of their time enacting the observation/feedback routine, but the majority of press they gave focused on content-neutral and often superficial feedback. Rhetorically, administrators were the instructional leaders of their campuses: but what did this mean? We demonstrated that administrators most often gave feedback

about basic instructional practices, such as expressing clear expectations for student learning, monitoring student engagement, or (less commonly) teachers' questioning strategies. In essence, most administrators were compliance monitors, attending to only the most basic of instructional practices. Our findings indicate that the current policies that mandate school administrators to spend multiple hours a day in classrooms are unlikely to contribute to teachers' development of ambitious instructional practices unless significant resources are invested to support administrator learning. We know that this is not easy (as our findings make that clear), and may also not be the most strategic use of administrator time or scarce district resources.

For Policy

Administrators are expected to manage the school building, evaluate teachers, work with parents, negotiate with unions, manage a staff, and many other tasks. Additionally, they are also expected to support teachers who have a wide variety of experience and expertise across multiple disciplines. In these four districts that expected school administrators to engage in the observation and feedback, less than 25% of teachers reported press that had the potential to support them in reorganizing their instructional practices. Furthermore, there is no evidence that the administrator press had impact on the quality of teachers' instruction, as evidenced by a lack of improvement in Instructional Band. In short, administrators spent a large percentage of their time observing teachers and providing feedback with no measurable impact on the quality of their instruction as measured by Instructional Band. Our findings strongly indicate that central district offices should change their priorities if we are to see improvement at school sites.

Our findings raise a number of questions. Future studies should aim to clarify the connections between the nature of teachers' instruction, the background and vision of high-quality instruction of the administrator, and the ways in which the observation and feedback routine is conceptualized and enacted. This would likely include qualitative observation to more fully document the observation/feedback routine as a whole: the lesson itself, the specific feedback the administrator gives to the teacher about the lesson, the decision-making processes the administrator uses to decide on that feedback, and how the routine plays into other instructional leadership activities on campus. In addition, other research suggests a positive relationship between a longer tenure in one school and positive outcomes (Coelli & Green, 2012); these findings imply the need for longitudinal studies. The current focus on school administrators as instructional leaders, however ambiguously defined (Rigby, 2014), calls for both researchers and district leadership to consider

appropriate and feasible expectations if we are to leverage expertise toward instructional improvement for all students.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported on in this paper was supported by the National Science Foundation under grants DRL-0830029 and ESI-0554535.

Notes

1. Unless the data or claim references principals specifically, we refer to "administrators" or "school leaders" rather than "principals" as many assistant principals are also expected to be instructional leaders and participate in observation and feedback routines.
2. Standardized tests need not be exclusively procedural, yet at this historical juncture they tend to be.
3. We also included these variables in early versions of our models, but they did not change the outcomes of our variables of interest.
4. Note that many of these analyses come from data from years of the project not included in this article.
5. Note that our data illustrate adequate classroom management. IQA coders reported that classroom management was not a barrier for instruction in any of the videos.
6. While the protocol asked specifically about principals, if the interviewee was observed by their assistant principal, we collected information about that individual. Since this distinction is difficult to determine across years, districts, schools, and individuals, we refer to the supervisor as the "administrator."
7. Note that we took a more-inclusive rather than strict approach to "what counts" as matching administrator press to instruction type. This is largely due to the nature of our data: We base our findings on annual interviews that cover a large amount of terrain, and therefore sometimes only scratch the surface with some questions. In light of this, and given the nuances that are impossible to discern with interviews, we think we are more likely to see trends in administrator press with a more-inclusive approach.
8. Note that coders often varied in their unitization, or where a code starts and ends, to some degree. This was expected, as we did not have discrete units of analysis. We therefore could not establish quantitative reliability scores using NVivo. Instead, we rely on the research of qualitative researchers such as Armstrong, Gosling, Weinman, and Marteau (1997), who argue that consensus on the theme is the relevant element for reliability rather than the exact "packaging" (p. 605).

9. We also conducted a series of analyses on teacher reports of frequency of administrator feedback, experience as an administrator, and whether or not principals taught mathematics prior to becoming an administrator. None of these analyses yielded statistically significant coefficients nor described the variance of the model in such a way that warranted inclusion.
10. To assess the accuracy of the final logistic regression model, we did a percent correctly predicted analysis to generate a goodness-of-fit measure. In total, the regression model correctly predicted the outcome approximately 80% of the time (215/271).
11. Prior to the analysis, we conducted a proportional odds test, which was not statistically significant.

References

Armstrong, D., Gosling, A., Weinman, J., & Marteau, T. (1997). The place of inter-rater reliability in qualitative research: An empirical study. *Sociology*, *31*, 597-606.

Ball, D. L., & Cohen, D. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes & L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3-32). San Francisco, CA: Jossey-Bass.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*, 389-407.

Barth, R. S. (1986). Principal centered professional development. *Theory Into Practice*, *25*, 156-160.

Blase, J., & Blase, J. (1999). Principals' instructional leadership and teacher development: Teachers' perspectives. *Educational Administration Quarterly*, *35*, 349-378.

Borko, H., Jacobs, J., Koellner, K., & Swackhamer, L. (2015). *Mathematics professional development: Improving teaching using the Problem-Solving Cycle and Leadership Preparation models*. New York, NY: Teachers College Press.

Boston, M. (2012). Assessing instructional quality in mathematics. *Elementary School Journal*, *113*, 76-104.

Boston, M., Henrick, E., Gibbons, L., Berebitsky, D., & Colby, G. (2016). Investigating how to support principals as instructional leaders in mathematics. *Journal of Research in Leadership Education*. Advance online publication. doi:10.1177/1942775116640254

Brinko, K. T. (1993). The practice of giving feedback to improve teaching: What is effective? *Journal of Higher Education*, *64*, 574-593.

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

Carver, C., Steele, M., & Herbel-Eisenmann, B. (2010). Instructional leadership = Principals + Algebra (− fear). *Journal of Staff Development*, *31*(5), 30-33.

Cobb, P., & Jackson, K. (2011a). Assessing the quality of the common core state standards for mathematics. *Educational Researcher*, *40*, 183-185.

Cobb, P., & Jackson, K. (2011b). Towards an empirically grounded theory of action for improving the quality of mathematics teaching at scale. *Mathematics Teacher Education & Development*, *13*(1), 6-33.

Cobb, P., & Jackson, K. (2012). Analyzing educational policies: A learning design perspective. *The Journal of the Learning Sciences*, *21*(4), 487-521.

Cobb, P., Jackson, K., Smith, T., Sorum, M., & Henrick, E. (2014). Design research with educational systems: Investigating and supporting improvements in the quality of mathematics teaching and learning at scale. In W. R. Penuel, B. J. Fishman, & B. H. Cheng (Eds.), *National Society for the Study of Education Yearbook on Design-Based Implementation Research*. New York: Teachers College Record.

Cobb, P., & Smith, T. (2008). The challenge of scale: Designing schools and districts as learning organizations for instructional improvement in mathematics. In T. Wood, B. Jaworski, K. Krainer, P. Sullivan, & D. Tirosh (Eds.), *International handbook of mathematics teacher education* (Vol. 3, pp. 231-254). Rotterdam, Netherlands: Sense.

Coburn, C. E., & Russell, J. L. (2008). District policy and teachers' social networks. *Educational Evaluation and Policy Analysis*, *30*, 203-235.

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

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, *12*, 311-329.

Cohen, D. K., & Hill, H. C. (2008). *Learning policy: When state education reform works*. New Haven, CT: Yale University Press.

Council of Chief State School Officers. (2015). *ISLLC 2015: Model policy standards for educational leaders*. Washington, DC: Author.

Darling-Hammond, L., Hightower, A. M., Husbands, J. L., LaFors, J. R., Young, V. M., & Christopher, C. (2003). *Building instructional quality: "Inside-out" and "outside-in." Perspectives on San Diego's school reform* (Research Report). Retrieved from https://depts.washington.edu/ctpmail/PDFs/Instructional Qual-09-2003.pdf

Darling-Hammond, L., Wise, A., & Pease, S. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, *53*, 285-328.

Elmore, R. F. (1996). Getting to scale with good educational practice. *Harvard Educational Review*, *66*(1), 1-27.

Elmore, R. F., & Burney, D. (1997). *Investing in teacher learning: Staff development and instructional improvement in Community School District #2, New York City*. New York, NY: National Commission on Teaching & America's Future, Teachers College.

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, *48*, 94-118.

Frank, K. A., Zhao, Y., & Borman, K. (2004). Social capital and the diffusion of innovations within organizations: The case of computer technology in schools. *Sociology of Education*, *77*, 148-171.

Gamoran, A., Anderson, C. W., Quiroz, P. A., Secada, W. G., Williams, T. & Ashmann, S. (Eds.). (2003). *Transforming teaching in math and science: How schools and districts can support change*. New York, NY: Teachers College Press.

Glanz, J. (2005). *What every principal should know about instructional leadership*. Thousand Oaks, CA: Corwin Press.

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.

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, *111*, 2055-2100.

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice*, *15*, 273-289.

Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, *9*, 157-191.

Halverson, R., Kelley, C., & Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. In W. K. Hoy & C. G. Miskel (Eds.), Educational administration, policy, and reform: Research and measurement (pp. 153-188). Greenwich, CT: Information Age.

Hiebert, J. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

Hiebert, J., Gallimore, R., Garnier, H. E., Givvin, K. B., Hollingsworth, H., Jacobs, J., & Stigler, J. W. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study*. Washington, DC: U.S. Department of Education.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol.1, pp. 271-405). Greenwich, CT: Information Age.

Hiebert, J., & Stigler, J. W. (2000). A proposal for improving classroom teaching: Lessons from the TIMSS video study. *Elementary School Journal*, *101*(1), 3-20.

Hilberg, R. S., Waxman, H. C., & Tharp, R. G. (2004). Introduction: Purposes and perspectives on classroom observation research. In H. C. Waxman, R. G. Tharp & R. S. Hilberg (Eds.), *Observational research in U.S. classrooms* (pp. 1-20). New York, NY: Cambridge University Press.

Hopkins, M., & Woulfin, S. L. (2015). School system (re)design: Developing educational infrastructures to support school leadership and teaching practice. *Journal of Educational Change*, *16*, 371-377.

Horn, I. S. (2010). Teaching replays, teaching rehearsals, and re-visions of practice: Learning from colleagues in a mathematics teacher community. *Teachers College Record*, *112*, 225-259.

Horn, I. S., Garner, B., Kane, B. D., & Brasel, J. (2016). A taxonomy of instructional learning opportunities in teachers' workgroup conversations. *Journal of Teacher Education*. Advance online publication. doi:10.1177/0022487116676315

Horn, I. S., Kane, B. D., & Wilson, J. (2015). Making sense of student performance data: Data use logics and mathematics teacher learning opportunities. *American Educational Research Journal*, *52*, 208-242.

Horn, I. S., Kane, B. D., Appelgate, M. A., Brasel, J., Garner, B., & Wilson, J. (2017). *Understanding access to expertise: An exploratory mixed methods approach*. Manuscript submitted for publication.

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

Horng, E., & Loeb, S. (2010). New thinking about instructional leadership. *Phi Delta Kappan*, *92*(3), 66-69.

Jackson, K., Garrison, A., Wilson, J., Gibbons, L., & Shahan, E. (2013). Exploring relationships between setting up complex tasks and opportunities to learn in concluding whole-class discussions in middle-grades mathematics instruction. *Journal for Research in Mathematics Education*, *44*, 646-682.

Kane, B. D. (2017). Facilitating teacher workgroups: Conceptually grounding local representations of practice. *Journal of Learning Sciences*. Manuscript submitted for publication.

Kazemi, E., Franke, M., & Lampert, M. (2009). *Developing pedagogies in teacher education to support novice teachers' ability to enact ambitious instruction*. Paper presented at the 32nd annual conference of the Mathematics Education Research Group of Australasia, Palmerston North, New Zealand.

Lampert, M., Beasley, H., Ghousseini, H., Kazemi, E., & Franke, M. (2010). Using designed instructional activities to enable novices to manage ambitious mathematics teaching. In M. K. Stein & L. Kucan (Eds.), *Instructional explanations in the disciplines* (pp. 129-141). New York, NY: Springer.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning: Executive Summary*. Minneapolis: University of Minnesota Center for Applied Research and Educational Improvement.

Lortie, D. C. (2009). *School principal: Managing in public*. Chicago, IL: University of Chicago Press.

Louis, K. S., 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.

Louis, K. S., Leithwood, K., Wahlstrom, K. L., & Anderson, S. E. (2010). *Investigating the links to improved student learning: Final report of research findings* (Learning from Leadership Project). St. Paul: University of Minnesota, Center for Applied Research and Educational Improvement.

Louis, K. S., Marks, H. M., & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal*, *33*, 757-798.

Matsumura, L. C., Garnier, H. E., Slater, S. C., & Boston, M. (2008). Toward measuring instructional interactions "at-scale." *Educational Assessment*, *13*, 267-300.

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

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, *83*, 340-365.

Murphy, J., Hallinger, P., & Heck, R. H. (2013). Leading via teacher evaluation: The case of the missing clothes? *Educational Researcher*, *42*, 349-354.

National Policy Board for Educational Administration. (2015). *Professional standards for educational leaders 2015*. Reston, VA: Author.

Nelson, B. S. (1997). *Lenses on learning: How administrators' ideas about mathematics, learning and teaching influence their approaches to action in an era of reform* (Center for the Development of Teaching Paper Series). Newton, MA: Education Development Center.

Newmann, F. M., King, M. B., & Youngs, P. (2000). Professional development that addresses school capacity: Lessons from urban elementary schools. *American Journal of Education*, *108*, 259-299.

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

O'Day, J., & Quick, H. E. (2009). Assessing instructional reform in San Diego: A theory-based approach. *Journal of Education for Students Placed at Risk*, *14*(1), 1-16.

Ovando, M., & Ramirez, A. (2007). Principals' instructional leadership within a teacher performance appraisal system: Enhancing students' academic success. *Journal of Personnel Evaluation in Education*, *20*, 85-110.

Penuel, W. R., Riel, M., Krause, A., & Frank, K. (2009). Analyzing teachers' professional interactions in a school as social capital: A social network approach. *Teachers College Record*, *111*, 124-163.

Protheroe, N. (2009). Using classroom walkthroughs to improve instruction. *Principal*, *88*(4), 30-34.

Remillard, J. T. (1999). Curriculum materials in mathematics education reform: A framework for examining teachers' curriculum development. *Curriculum Inquiry*, *29*, 315-342.

Rigby, J. G. (2014). Three logics of instructional leadership. *Educational Administration Quarterly*, *50*(4), 610-644.

Rigby, J. G., Andrews-Larson, C., & Chen, I. (2014). *Shifting teacher views of mathematics instruction and student struggle: A mixed-methods analysis*. Paper presented at the American Educational Research Association, Philadelphia, PA.

Robinson, V. M., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, *44*, 635-674.

Sassi, A. M., & Nelson, B. S. (1999). *Learning to see anew: How facilitator moves can reframe attention when administrators look at reformed mathematics classrooms*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Scheeler, M. C., Ruhl, K. L., & McAfee, J. K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, *27*, 396-407.

Schoenfeld, A. H. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational Researcher*, *31*(1), 13-25.

Shin, S., & Slater, C. L. (2010). Principal leadership and mathematics achievement: An international comparative study. *School Leadership & Management*, *30*, 317-334.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*, 153-189. doi:10.3102/0034654307313795

Spillane, J. P., & Callahan, K. A. (2000). Implementing state standards for science education: What district policy makers make out of the hoopla. *Journal of Research in Science Teaching*, *37*, 401-425.

Spillane, J. P., Diamond, J. B., Burch, P., Hallett, T., Jita, L., & Zoltners, J. (2002). Managing in the middle: School leaders and the enactment of accountability policy. *Educational Policy*, *16*, 731-762.

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning*, *10*, 313-340.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*, 455-488.

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, *2*(1), 50-80.

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

Sztajn, P., Borko, H., & Smith, T. (in press). Research on mathematics professional development. In J. Cai (Ed.), *Compendium for research in mathematics education*. Reston, VA: National Council of Teachers of Mathematics.

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

Thompson, A. G., Philipp, R. A., Thompson, P. W., & Boyd, B. A. (1994). Calculational and conceptual orientations in teaching mathematics. In D. B. Aichele & A. F. Coxford (Eds.), *Professional development for teachers of mathematics* (pp. 79-92). Reston, VA: National Council of Teachers of Mathematics.

Tourangeau, R., Rips, L., & Rasinski, K. (2000). *The psychology of survey response*. New York, NY: Cambridge University Press.

Van Houten, R. (1980). *Learning through feedback: A systematic approach for improving academic performance*. New York, NY: Human Sciences Press.

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.

Wolcott, H. (1973). *The man in the principal's office: An ethnography*. Prospect Heights, IL: Waveland Press.

Zepeda, S. J. (2007). *The principal as instructional leader: A handbook for supervisors* (2nd ed.). Larchmont, NY: Eye on Education.

Author Biographies

Jessica G. Rigby is an assistant professor in the Education Policy, Organizations, and Leadership in P12 Systems in the College of Education at the University of Washington. Before coming to the University of Washington she received her PhD from the University of California, Berkeley, and did a 2-year postdoctoral fellowship at Vanderbilt University on the MIST project. She uses lenses from organizational sociology to understand the role of school and district leaders in the implementation of policy and classroom instruction.

Adrian Larbi-Cherif is a doctoral candidate in math education at Vanderbilt University's Peabody College. His research focuses on understanding how instructional leadership and the processes of policy implementation influence math teacher's development of ambitious instructional practices.

Brooks A. Rosenquist is a doctoral candidate in the Department of Education Leadership, Policy, and Organizations at Vanderbilt University. His research interests center on data-driven school improvement, researcher–district partnerships in STEM learning, and policies and structures to improve student engagement and motivation.

Charlotte J. Sharpe is a doctoral candidate in math education at Vanderbilt University's Peabody College. Her research focuses on designs and methodologies for supporting and measuring math teachers' learning of ambitious instructional practices.

Paul Cobb is a professor of mathematics education at Vanderbilt University, Department of Teaching and Learning. His research focuses on instructional design, issues of equity in mathematics teaching and learning, and the improvement of mathematics teaching on a large scale.

Thomas Smith is a professor and the dean of the Graduate School of Education at the University of California, Riverside. His research focuses on identifying supports and barriers to improving teaching and learning at the scale of large urban school districts.