American Educational Research Journal
June 2015, Vol. 52, No. 3, pp. 475–514
DOI: 10.3102/0002831215585562

Teacher Collaboration in Instructional Teams and Student Achievement

Matthew Ronfeldt
Susanna Owens Farmer
Kiel McQueen
University of Michigan
Jason A. Grissom
Vanderbilt University

This study draws upon survey and administrative data on over 9,000 teachers in 336 Miami-Dade County public schools over 2 years to investigate the kinds of collaborations that exist in instructional teams across the district and whether these collaborations predict student achievement. While different kinds of teachers and schools report different collaboration quality, we find average collaboration quality is related to student achievement. Teachers and schools that engage in better quality collaboration have better achievement gains in math and reading. Moreover, teachers improve at greater rates when they work in schools with better collaboration quality. These results support policy efforts to improve student achievement by promoting teacher collaboration about instruction in teams.

KEYWORDS: professional development, professional learning communities, student achievement, teacher collaboration, teacher and school quality

Teaching in the United States historically has been isolated work (Lortie, 1975), but in recent years, reformers have pushed to transform schools into places where teachers work collectively on instruction. Policymakers have called for the creation of school-based professional learning communities (PLCs; National Staff Development Council, 2001) and for organizational structures that promote regular opportunities for teachers to collaborate with teams of colleagues (Carroll, 2007; Hamilton et al., 2009). These initiatives are based on the assumption that collaboration enables teachers to strengthen their instruction, thus improving learning outcomes for students (Bryk, Camburn, & Louis, 1999; Darling-Hammond & Richardson, 2009). Although calls for collaboration have become widespread, few large-scale studies have investigated how these calls have been taken up in practice. As collaboration becomes increasingly common (MetLife Foundation, 2009), it is critical that scholars understand this key facet of present-day contexts of teaching.

Researchers are only beginning to understand how teacher collaboration affects student achievement. There is some evidence that schools characterized by higher levels of collaboration also have higher levels of student achievement (Goddard, Goddard, & Tschannen-Moran, 2007; Goddard, Miller, Larson, & Goddard, 2010). However, we do not know whether this relationship is causal, nor do we understand the mechanisms (e.g., changes in individual teachers' practice or in school culture) by which instructional collaboration impacts students' learning. Furthermore, we do not know whether certain kinds of collaboration (e.g., focused on specific instructional domains) have a stronger influence on student achievement than other kinds do.

This article contributes to the literature on instructional collaboration in at least four ways. First, we provide a descriptive account of the kinds and quality of collaboration that exist in instructional teams across a large, urban district, and we show how collaborations vary for different kinds of teachers and schools. This district-level view of existing collaborations—which, to our knowledge, is the first of its kind—reveals which instructional domains

---

Matthew Ronfeldt is an assistant professor of educational studies at the University of Michigan School of Education, 610 East University Ave, Ann Arbor, MI 48109; e-mail: *ronfeldt@umich.edu*. His scholarship seeks to understand how to improve teaching quality, particularly in urban and underserved settings. His research sits at the intersection of educational practice and policy and focuses on teacher preparation, teacher retention, teacher induction, and the evaluation of teachers and preparation programs.

Susanna Owens Farmer is a doctoral candidate in educational studies at the University of Michigan School of Education, 610 East University Ave, Ann Arbor, MI 48109; e-mail: *susanna.farmer@umich.edu*. Her research focuses on resources for elementary teaching (e.g., curriculum materials) and the ways they contribute to the interactions teachers have with students around content. She is also interested in understanding the kinds of structures and supports that enable practicing teachers to participate in collective work on teaching.

Kiel McQueen is a doctoral candidate in educational studies at the University of Michigan School of Education, 610 East University Ave, Ann Arbor, MI 48109; e-mail: *kielmcqu@umich.edu*. His research focuses on supports for beginning teachers (e.g., induction) and the ways that evaluation and instructional coaching might improve teachers' practice and retention, particularly in urban and underserved contexts.

Jason A. Grissom is an associate professor of public policy and education in the Department of Leadership, Policy, and Organizations at the Peabody College of Education and Human Development, Vanderbilt University, PMB 414, Nashville, TN 37203; e-mail: *jason.grissom@vanderbilt.edu*. His research utilizes large-scale administrative and survey data to investigate teacher and school leader effectiveness, evaluation, retention, and mobility, with a particular interest in the strategies school leaders employ to improve teaching and learning in their schools.

(e.g., assessment, teaching strategies) receive more and less attention in teachers' instructional teams as well as which kinds of collaboration teachers perceive to be more and less helpful. Second, we test whether certain kinds of collaboration are more strongly related to student achievement than others. Specifically, we examine how both the quality and content of collaboration are related to student outcomes. Third, we investigate possible mechanisms by which collaboration influences student achievement. In particular, we make progress in differentiating the effects of working in a more collaborative school from being a more collaborative individual. Finally, we provide suggestive evidence of a causal relationship between collaboration and student achievement by controlling for a comprehensive set of school and teacher characteristics that might explain observed relationships and by testing various threats to a causal interpretation of the estimates.

Literature Review

Before describing our study, we review existing literature related to instructional collaboration, including studies that provide descriptive information about collaboration and professional communities in schools, studies that link collaboration to student achievement, and studies that examine features of collaboration related to student achievement gains. This review of selected literature offers background and motivation for our study.

The Landscape of Collaboration: Variation by Teacher and School Characteristics

Although a number of studies have described the nature of collaboration occurring among particular groups of teachers (Horn & Little, 2009; Levine & Marcus, 2010; Little, 2002, 2003; Strahan, 2003), little is known about the current social-institutional landscape of teacher collaboration across teachers and schools. The few existing large-scale descriptions suggest there is wide variation in teachers' collaborations. Drawing upon elementary teacher surveys in a large, urban district, Goddard et al. (2007) found significant differences between schools in the extent to which teachers reported collaborating on decisions about instructional improvement, such as selecting instructional methods and evaluating curriculum. A study of a nationally representative sample of K–12 public school teachers also suggested that teacher collaboration varies widely (MetLife Foundation, 2009). While virtually all respondents reported participating in collaborations, they collaborated from less than 30 minutes per week (12%) to more than 3 hours per week (24%). Survey results also suggested that some kinds of collaboration are more common than others. For example, many teachers reported meeting frequently with colleagues to discuss ways to increase achievement and to examine student work (75% and 68%, respectively), but far fewer (22%) reported observing colleagues to provide instructional feedback.

These studies provide important initial descriptions of the nature of teachers' collaborations, but they focus primarily on the amount or extensiveness of certain kinds of collaboration. In an era where collaboration among faculty is increasingly promoted as a means for improving schools, it is critical to understand not only the *extent* to which teachers collaborate about certain topics but also the degree to which different kinds of collaboration are *useful* in supporting their practice. Thus, we extend prior work by focusing on both the extensiveness and helpfulness of collaboration that teachers report in their instructional teams.

Existing literature also suggests that school and teacher characteristics likely influence the ways in which teachers interact with their colleagues. For example, Louis, Marks, and Kruse (1996) found that elementary schools and schools with more female teachers were more likely to have higher levels of "professional community," a construct that encompassed collaboration and other elements, such as de-privatized practice and shared norms and values. In a study of Chicago elementary schools, Bryk et al. (1999) identified several factors that facilitated professional community, including small school size, strong principal leadership, and social trust among faculty. In addition, Bryk et al. found that certain kinds of teachers—those who were African American, actively involved in school committees and activities, and more experienced—were more likely to report higher levels of professional community in their schools.

Linking Collaboration and Student Achievement

Although multiple studies have shown school climate to predict student achievement (Kraft & Papay, 2014; Lee & Smith, 1996; Louis & Marks, 1998), these studies do not disentangle the effects of teacher collaboration from other dimensions of climate. Goddard et al.'s (2007) study was among the first large-scale studies to demonstrate a link specifically between teacher collaboration and student achievement. This study found that elementary schools with higher levels of collaboration also had higher levels of student achievement, even after controlling for a set of student-level variables (i.e., gender, race, socioeconomic status [SES], and prior achievement) and school-level variables (i.e., school size, school-wide SES, and proportion minority students). In a follow-up study, Goddard et al. (2010) tested the link between student achievement and a broader measure of collaboration, which included teachers' ratings on three different scales: frequency of collaboration on instruction, the extent to which teachers collaborate on instructional policy (used in their 2007 study), and teachers' participation in formal collaboration structures. Using structural equation models, the authors reported a direct effect of teacher collaboration on student achievement and an indirect effect of principals' instructional leadership on student achievement mediated by collaboration.

While providing critical evidence for a relationship between teacher collaboration and student achievement, both studies also had important limitations. Namely, these studies used collaboration measures that were (1) domain-general, without disentangling the effects of collaboration about different instructional topics; (2) based on measures for the amount (i.e., frequency or extensiveness) of collaboration without attending to other dimensions of collaboration quality; and (3) aggregate, school-level measures, which essentially ignored within-school differences in collaboration and their effects. Regarding the latter, we find the vast majority of variation in collaboration is within, not between, schools, suggesting the need for attention to differences in collaboration even among teachers in the same school environment.

Looking Beyond Frequency: Linking Other Collaboration Features to Student Achievement

Some studies—mainly case studies and quasi-experimental intervention studies—have gone beyond frequency to consider other collaboration features that support teacher and student learning. A common theme in this literature is that all collaborations are not equal—or equally productive. In case studies of discourse patterns among teacher teams, Little (2003) and Horn and Little (2009) argued that the nature of teachers' conversations with colleagues serve to either open up or close down opportunities for teacher learning, depending in part on the conversational norms that guide these interactions. Two major implications are (1) the quality of collaboration can vary across instructional teams, even among teams that collaborate regularly and are highly committed to students' learning, and (2) the quality of collaboration likely influences the degree to which teams are able to change teachers' practice and ultimately improve student learning.

Several studies have attempted to identify features of "quality" collaboration, and some have sought to link specific kinds of collaboration to student learning outcomes. Our review of the literature indicates two kinds of collaboration likely promote gains in students' learning: (1) collaboration focused on analyzing student data and developing instructional responses and (2) collaboration focused on curriculum and instructional decision-making. We elaborate below.

Collaboration focused on analyzing student data and developing instructional responses. One way researchers have sought to identify features of collaboration that may support student achievement is by examining the kinds of collaboration that exist in exceptionally high-performing schools. Strahan (2003), for example, conducted case studies of three elementary schools that exhibited higher-than-expected levels of student achievement in order to identify features of the professional cultures of these schools. He observed that teachers in these schools focused their collaborations on identifying students' learning needs and then designing ways to address these needs. Strahan characterized these collaborations as "data-directed dialogue" (p. 143) because they were informed by data from both formal assessments and informal observations of students' learning. Similarly, in their review of the literature on PLCs, Vescio, Ross, and Adams (2008) concluded that the most effective PLCs were those characterized by "collaboration with a clear and persistent focus on data about student learning" (p. 89).

The implications of these studies—that collaboration around student data may be useful for raising student achievement—align well with the current emphasis on data-driven decision-making in schools (Hamilton, Halverson, Jackson, Mandinach, Supovitz, & Wayman, 2009). As Coburn and Turner (2011) point out, however, the impact of data-use processes hinges upon how teachers make sense of these data, which in turn is shaped by the organizational and political contexts in which teachers work, including institutional norms, routines, and leadership. Thus, the degree to which teachers' collaborations around student data positively influence student learning will depend upon the nature of these collaborations and their organizational contexts.

Saunders, Goldenberg, and Gallimore (2009) and Gallimore, Ermeling, Saunders, and Goldenberg (2009) describe quasi-experimental research on whether supporting teacher teams to engage in inquiry around student data increases student achievement. Teachers in the treatment schools worked in grade-level teams to identify students' learning needs and develop ways to address them. The nine treatment schools demonstrated significantly higher student achievement gains than the six control schools, but only in the second phase of the intervention, when collaboration in treatment schools was more frequent, explicitly structured around inquiry-focused protocols, and led by trained members of school instructional leadership teams. These studies provide initial evidence that inquiry-focused collaboration around student work may contribute to increases in student learning, but it is difficult to disentangle the effects of the collaboration from the effects of other components of the larger intervention. Also, because the research focused on carefully designed collaborations that likely differ from the kinds that emerge naturally across school settings, we do not know whether more typical forms of teacher collaboration improve student achievement.

Collaboration focused on curriculum and instructional decision-making. A number of other studies indicate that collaboration focused on instructional domains, like curriculum and teaching strategies, may also be promising. In an evaluation in Cincinnati, Supovitz (2002) found that teacher teaming did not influence student achievement on average. However, teams that maintained a high level of "group instructional practice"—preparing together for instruction, co-teaching, observing one another, and grouping students flexibly for particular instructional purposes—had better student achievement. These results indicate that collaboration focused on instructional planning and enactment may improve student outcomes. Similarly, Goddard et al. (2007), described earlier, found collaboration focused on curriculum and instructional decision-making to predict school-level achievement.

Our study extends prior research on features of collaboration and their relationship with student achievement in a number of ways. First, we develop novel measures for collaboration quality in order to build upon prior literature suggesting that the quality of collaboration matters. Rather than focus on specific structures or norms that create conditions for "quality" collaboration to occur, the approach taken by most prior research, we measure teachers' experiences of quality more directly by constructing measures based upon their reports of the helpfulness and extensiveness of collaboration in different instructional domains. Second, we extend prior literature about the content of teacher collaboration, which tends to investigate multiple domains of collaboration in a single measure. We instead conduct exploratory factor analysis to identify a range of common topics taken up during collaborations in instructional teams and then test whether the quality of collaboration in each of these domains predicts student achievement gains. To our knowledge, no prior research has attempted to disentangle the relative impacts of collaboration about different instructional domains.

Summary and Research Foci

Existing literature provides evidence that teacher collaboration is associated with student achievement. However, it does not yet provide a clear understanding about (1) the types of collaboration that are common across schools and whether these vary by teacher and school characteristics, (2) the collaboration features that are most strongly associated with student achievement gains, or (3) how within-school differences in collaboration may be related to student achievement. To our knowledge, all prior large-scale research has studied collaboration as a school-level phenomenon. Furthermore, existing literature has done little to determine whether observed relationships between collaboration and achievement are causal in nature.

This study, then, begins with a descriptive analysis of teacher collaboration, asking the following: What kinds of instructional collaborations exist in this large urban district? Do teachers perceive collaboration in certain instructional domains to be more extensive or helpful than others? How much variation in collaboration quality exists within and between schools? We then investigate differences in teachers' reports of collaboration by school and teacher characteristics. Finally, we test whether the quality of teachers' collaboration is associated with achievement gains or how quickly teachers improve with experience.

Data and Methods

Data

This study draws on extensive teacher survey and administrative data from the Miami-Dade County Public School System (MDCPS). MDCPS is the fourth largest school district in the United States, enrolling more than 350,000 students. The sample comprises over 9,000 teacher observations across all MDCPS schools over 2 academic years (2010–2011 and 2011–2012, which we refer to as 2011 and 2012, respectively). Our surveys collected comprehensive information about the collaborations that teachers had in their instructional teams. We then linked the survey information to district administrative data on the teachers, their schools, and the students with whom they worked, including test score data.

Survey data come from a larger study of school leaders in MDCPS (see Grissom & Loeb, 2011; Grissom, Loeb, & Master, 2013). In 2011, all MDCPS teachers were surveyed; in 2012, a random sample of teachers in each school was surveyed. We did not survey teachers who worked in charter schools. Response rates were 36% in 2011 and 39% in 2012, resulting in 8,058 and 1,248 respondents, respectively. Surveys were conducted online in May and June each year. Each teacher received an individual link through his or her district e-mail account that allowed responses to be linked to personnel and other administrative records. Teachers responded to questions covering a range of topics, including school climate, leadership, and supports for teaching. Our analyses focus on a set of questions about the extensiveness and helpfulness of different kinds of collaboration that teachers had in their instructional teams.

District administrative records provided data about students, teachers, and schools. Personnel files included data on teacher demographics (gender, race/ethnicity), degree information, as well as job duration, place (school), and type (instructional, special education). School administrative files provided enrollment and aggregate student demographic data, and student administrative files included individual demographic and achievement data.

Sample

The upper bound of our analytic sample was 336 schools and 7,881 teachers for reading models (6,682 for math). However, our analytic sample varied, largely depending upon whether we were investigating teacher- or school-level phenomena and which years were included. When examining school-level value-added (described below) as a function of school-level collaboration, our sample included 336 schools in 2011 for which we could calculate both school-level value-added and collaboration scores. When examining teacher-level value-added as a function of both teacher- and school-level value-added, our sample was restricted to those teachers who responded to the survey and had school- and teacher-level value-added information in 2011 and/or 2012 (667 teachers in reading, 544 in math). To investigate whether individual teachers improved at faster rates in schools with better collaboration, we assumed school-level collaboration measured in 2011 was constant from 2010–2012 (a limitation we discuss in "Methods"). Here, our analytic sample included all teachers with at least one value-added score from 2010–2012 in one of the focal 336 schools (7,881 reading teachers; 6,682 math); school-level collaboration measures were available even for teachers who did not complete surveys.

Table 1 summarizes descriptive statistics for the schools in our sample. Over half (55%) of these schools were elementary level. About three quarters (76%) of students qualified for free or reduced priced lunch and almost one half (48%) of students were classified as limited-English proficient (LEP). An average of over 90% of students were either Hispanic (58%) or Black (33%). Spoken language also varied within this district, with Spanish being the first language for 48% of students and English for 44%. Additionally, at least one out of five students qualified either for Exceptional Student Education (ESE) or gifted student services.

Table 2 summarizes descriptive statistics for (1) all teachers in the district in 2011 and 2012 for whom either math or reading (or both) value-added scores were available (N = 8,602) and (2) the subset of (1) for whom teacher-level collaboration factor scores were also available (N = 1,109).[1] Similar to national patterns, the majority of teachers in the district were female (83%). Unlike national patterns, most teachers in this district were non-White (88%). Almost half (48%) of teachers in the district were Hispanic and over a quarter were Black (28%). The majority of teachers (55%) held a BA as their highest degree, although over one third completed MA degrees (37%). The district classified the majority (82%) of teachers for whom it was possible to construct value-added scores as having "instructional" jobs. We were initially surprised that the proportion of teachers classified in "instructional" positions was not higher; however, we later discovered that 15% of teachers with value-added scores were classified as having ESE designations, which included instructional work with special education, gifted, and other exceptional student groups. Finally, participants had been employed in the district as teachers for almost 11 years, on average, including 27% that had worked in this capacity for 15 or more years.

We were able to construct collaboration measures for only a subset of teachers who responded to the survey (Table 2, right columns). On most characteristics, the subsample with collaboration measures was statistically similar to the broader sample. However, teachers in the subsample were more likely to be White ($p < .001$), less likely to by Hispanic ($p < .01$), more likely to have earned a specialist (non-PhD/MA) degree ($p < .001$),

Table 1
School Characteristics

Variable	Schools (2011) With Collaboration Factors and Value-Added (N = 336)	
	M	*SD*
Math value-added scores	–0.02	1.16
Reading value-added scores	0.07	0.73
Elementary	0.55	0.49
Middle school	0.18	0.38
PreK–8	0.09	0.29
Grades 6–12	0.03	0.16
High schools	0.14	0.35
Female	0.48	0.09
Free/reduced priced lunch	0.76	0.20
White	0.07	0.10
Black	0.33	0.34
Hispanic	0.58	0.32
Asian	0.01	0.12
Other race/ethnicity	0.01	0.02
English speaking	0.44	0.24
Spanish speaking	0.48	0.28
Other language	0.08	0.13
Limited English Proficient (LEP)	0.48	0.23
Exceptional Student Education (ESE)	0.12	0.11
Gifted, non-ESE	0.10	0.08
Average absences	9.03	5.69
Average suspensions	0.75	1.51
Enrollment per 100	9.15	6.65

Note. School-level value-added means are nonzero because they were constructed using our full sample of schools, including many charter schools not included in our survey sample. "Other race/ethnicity" includes students who identified as being Native American or multiracial/multi-ethnic.

and more experienced ($p < .01$). These differences raise some concern over survey nonresponse bias. Namely, if participating teachers differ on characteristics that also predict value-added then differences between populations might account for observed relationships between collaboration and value-added. Of the potential characteristics, only teaching experience has been shown to predict value-added. Differentially stronger response among more experienced teachers then could account for observed positive relationships between collaboration and value-added so long as experienced teachers also report better quality collaboration. However, we find years of experience to be unrelated to collaboration measures (see Table 2).

Table 2
Teacher Characteristics

Variable	Teachers (2011 and 2012) With Value-Added (*N* = 8,602) *M*	*SD*	Teachers (2011 and 2012) With Value-Added and Collaboration Factors (*N* = 1,109) *M*	*SD*
Female	0.83	0.38	0.83	0.37
White	0.22	0.42	0.27	0.45
Black	0.28	0.45	0.27	0.45
Hispanic	0.48	0.50	0.44	0.50
Asian	0.01	0.12	0.01	0.12
Native American	0.01	0.07	0.00	0.04
Other race/ethnicity	0.00	0.01	0.00	0.00
AA degree	0.00	0.04	0.00	0.03
BA degree	0.55	0.50	0.49	0.50
PhD degree	0.01	0.11	0.02	0.13
MA degree	0.37	0.48	0.39	0.49
Special degree	0.06	0.23	0.09	0.29
NA degree	0.01	0.08	0.01	0.07
Job: instruction	0.82	0.38	0.83	0.38
Job: ESE[a]	0.15	0.36	0.14	0.35
Job: vocational	0.00	0.03	0.00	0.00
Job: adult education	0.00	0.06	0.01	0.08
Job: other	0.02	0.14	0.02	0.15
In-district teaching experience[b]	10.78	8.37	11.54	8.64
Proportion 15-plus years	0.27	0.45	0.32	0.47

[a]This refers to educators who teach students who receive ESE services.
[b]This is number of years that an individual has been employed as a classroom teacher in the district. This does not include years of teaching outside of the district.

Additionally, the positive relationships between collaboration and value-added persist even in models that control for teacher characteristics. Thus, we doubt nonresponse bias explains this study's main findings.

Measures

Most surveyed teachers (84%) responded affirmatively to a question asking if they were a part of "a team or group of colleagues that works together on instruction." Because this study is focused on the relationships between the quality of collaboration that teachers have in their instructional teams and their student achievement gains, individuals who responded negatively to this item were excluded. Individuals who responded affirmatively were asked to respond to follow-up questions about the "extensiveness" and "helpfulness" of the collaborations with their instructional teams (see Tables 3 and 4). We applied exploratory factor analysis to these survey questions to examine their underlying structure and to identify latent constructs. These items loaded together with a high level of internal consistency (Cronbach's alpha = 0.93).

Domain-general collaboration factor. As shown in Table 3, all items loaded positively and strongly (0.6 to 0.8) on the first, unrotated factor; this factor had an eigenvalue of 7.9 and explained almost half of the variance in the items.[2] We decided to extract this unrotated factor as a measure for general collaboration across instructional domains. Because items about collaboration in different instructional domains did not load separately from one another, this factor reflects domain-general collaboration—hence our use of "general" to refer to this factor. We also constructed a composite measure by calculating the average of standardized survey items for each teacher. However, this composite measure was virtually indistinguishable from the general factor ($r = 0.998$), so we report only on the factor score.

Domain-specific collaboration factors. Though the first factor explained the vast majority of variation, a scree plot and the Kaiser criterion both suggested that it would also be reasonable to retain three factors, all with eigenvalues greater than 1: 7.9 for Factor 1, 1.5 for Factor 2, and 1.4 for Factor 3. To help with the interpretation of these extracted factors, we applied varimax rotation. Because we were interested in identifying conceptually distinct forms of collaboration, we deemed orthogonal rotation most appropriate. For comparison, we also rotated factors using nonorthogonal methods (oblimin). Results were similar so we report only on orthogonal ones. Table 3 summarizes these factors before and after rotation. The rotated factors loaded most strongly on the following sets of items:

- *Collaboration About Instructional Strategies and Curriculum*: Variables loading most strongly on this factor focus on collaboration about pedagogical and curricular approaches/strategies, including coordinating curriculum across classrooms, developing instructional strategies, and developing aligned materials (see Table 3, rows 1–2, 9–11).
- *Collaboration About Students*: Variables loading most strongly on this factor included collaboration about instructional topics/strategies focused on students, including discussing the needs of specific students, reviewing classroom work, and addressing student discipline/classroom management issues (see Table 3, rows 4, 7, 8, 12, 15, and 16).
- *Collaboration About Assessment*: Items loading on this factor focused on collaboration about assessments, including reviewing state test results and formative assessments (see Table 3, rows 5, 6, 13, and 14).

Table 3
Comparing Factor Loadings Before and After Rotations

	Rotated			Un-Rotated		
Variable	Factor1	Factor2	Factor3	Factor1	Factor2	Factor3
Extent developing curriculum and/or materials	0.79	0.18	0.17	0.68	–0.27	–0.38
Extent developing instructional strategies	0.64	0.31	0.33	0.75	–0.16	–0.16
Extent coordinating curriculum/ instruction	0.77	0.15	0.22	0.68	–0.31	–0.33
Extent addressing classroom management	0.21	0.73	0.09	0.61	0.46	–0.07
Extent reviewing state test results	0.12	0.14	0.85	0.59	–0.26	0.57
Extent reviewing formative assessments	0.26	0.18	0.78	0.68	–0.27	0.43
Extent reviewing students' classroom work	0.23	0.69	0.26	0.68	0.36	0.05
Extent discussing the specific student needs	0.15	0.74	0.26	0.67	0.43	0.10
Helpfulness developing curriculum/materials	0.80	0.24	0.19	0.74	–0.24	–0.37
Helpfulness developing instructional strategies	0.66	0.37	0.32	0.79	–0.12	–0.18
Helpfulness coordinating curriculum/instruction	0.78	0.23	0.23	0.74	–0.25	–0.33
Helpfulness addressing classroom management	0.30	0.74	0.13	0.69	0.42	–0.09
Helpfulness reviewing state test results	0.24	0.22	0.81	0.70	–0.23	0.47
Helpfulness reviewing formative assessments	0.35	0.28	0.73	0.76	–0.20	0.33
Helpfulness reviewing students' class work	0.29	0.70	0.26	0.73	0.34	0.01
Helpfulness discussing specific student needs	0.24	0.73	0.27	0.71	0.39	0.05

Given their strong correlations with the domain-general factor, we do not include domain-specific factors in models with this general factor. Instead, we first run models on this general factor to test whether domain-general collaboration predicts students' achievement; we then replace the domain-general factor with domain-specific factors to test whether collaboration

Table 4
Mean Responses to Survey Questions About Collaboration in Instructional Teams

Variable	Observations	*M*	*SD*	Min	Max
When you met with [your] instructional team, to what extent were the following covered . . . (1 = *not at all*; 2 = *a little*; 3 = *in some depth*; 4 = *in substantial depth*)					
Reviewing formative assessments	4,673	3.06	0.88	1	4
Developing instructional strategies	4,669	3.06	0.83	1	4
Reviewing state test results	4,664	3.02	0.94	1	4
Discussing the needs of specific students	4,680	3.01	0.88	1	4
Coordinating curriculum and/or instruction across classrooms	4,667	2.95	0.91	1	4
Developing curriculum and/or materials	4,664	2.94	0.92	1	4
Addressing classroom management/discipline issues	4,658	2.70	0.94	1	4
Reviewing students' classroom work	4,657	2.51	0.97	1	4
When you met with your instructional team, how helpful did you find each of the following activities . . . (1 = *not at helpful*; 2 = *a little helpful*; 3 = *helpful*; 4 = *very helpful*; 5 = *essential*)					
Developing instructional strategies	4,554	3.55	1.08	1	5
Discussing the needs of specific students	4,568	3.49	1.13	1	5
Developing curriculum and/or materials	4,532	3.45	1.14	1	5
Coordinating curriculum and/or instruction across classrooms	4,532	3.43	1.16	1	5
Reviewing formative assessments	4,542	3.43	1.14	1	5
Reviewing state test results	4,516	3.34	1.21	1	5
Addressing classroom management/discipline issues	4,497	3.13	1.20	1	5
Reviewing students' classroom work	4,473	2.97	1.21	1	5

about some instructional domains are more predictive of achievement than others.

Throughout our analysis, we interpret the four factors described above as measures of the "quality" of collaboration reported by teachers. We do so because we make the assumption that collaboration viewed as both extensive and helpful is of better quality. That is, when individuals experience high-quality collaboration, we assume they are more likely to rate its extensiveness and helpfulness as high; in turn, these same individuals should score higher on our collaboration factors that combine measures of both helpfulness and extensiveness. We acknowledge there are likely other dimensions of collaboration, beyond extensiveness and helpfulness, that contribute to its quality. While we may not measure all dimensions of quality, both extensiveness and helpfulness are probably necessary. Extensive collaboration that is unhelpful, like helpful collaboration that is not extensive, will probably not qualify as being "high quality." Combining information on helpfulness and extensiveness is an improvement over most prior research, which has focused on collaboration frequency. Even so, future research should further interrogate the psychometric properties involved in measuring collaboration quality.

Another possible limitation of our measures for collaboration is that they are based upon self-reported information. It is likely, for example, that some individuals will report poor quality collaboration to be of high quality (or vice versa) due to various psychological biases, subjective factors, and constraints of memory. While there may be some features that make for "better quality" collaboration for most individuals most of the time, we suspect the experience for a given individual will always depend to some degree on that individual's background, expectations, and perceptions; that is, quality will always remain subjective and contingent to some degree. For example, some forms of collaboration may be helpful to some but unhelpful to others, perhaps due to an individual's personal and professional needs. Subjective ratings of quality are likely, then, a better way to signal whether collaboration is responsive to these kinds of person-specific needs. In fact, a contribution of our work is that we include both teacher-level and aggregate, school-level measures for quality as a way to disentangle an individual's experience of collaboration from the average experiences of teachers throughout the school. If we assume that the quality of collaboration will depend upon both objective and subjective components, then the school-level measure should be sensitive to the former, and the teacher-level measure should be sensitive to the latter.

To construct measures for average collaboration quality at the school level, we aggregated teacher-level scores by school. Because only a fraction of teachers in each school were sampled in the 2012 survey, the number of respondents was quite low. Thus, we only constructed school-level scores in 2011. In this year, the proportions of teachers who responded to surveys varied by school, ranging from a minimum response rate of 15% to a maximum of 82%. Aggregate collaboration scores for schools with lower response rates are more prone to bias, so in our school-level regression analyses, we used response rates as probability weights to adjust estimates to weight more heavily schools with greater response rates.

Value-added measures. We constructed teacher-level and school-level value-added to student achievement scores based upon district test score information in both reading and mathematics. The test score data include Florida Comprehensive Assessment Test math and reading scores in Grades 3 through 10. We standardized test scores to have a mean of 0 and standard deviation of 1 within each grade and year. To construct teacher-level value-added in math and reading, we used a common approach of regressing students' test scores in math and reading (in separate equations) on their prior scores in both subjects, a series of student-level characteristics, classroom-level characteristics, school-level characteristics, grade-level indicators, and year indicators. Student characteristics included race/ethnicity, gender, free/reduced price lunch status, English proficiency status, number of absences in prior year, and number of suspension days in prior year. To construct classroom and school characteristics, we aggregated these student-level characteristics at the classroom and school levels, respectively; in addition, for school characteristics, we added enrollment. Each regression included a teacher-by-year fixed effect, which captured the difference between the average score of students in a teacher's classroom and the score that would be predicted for the students given their background characteristics. The estimates of these fixed effects were the value-added scores. We used a similar approach to construct school value-added measures, substituting school-by-year fixed effects for teacher-by-year fixed effects. Because they would be absorbed by school-by-year fixed effects, we did not include school-level covariates. In this case, coefficients on school-by-year fixed effects served as our school-level value-added estimates. For more details on our approach to constructing value-added, see Loeb, Kalogrides, and Beteille (2012).

Analytic Strategy

Do different kinds of schools have different kinds of instructional collaboration? To answer this research question, we ran a series of OLS regression models estimating a different school-level collaboration factor as a function of school characteristics. Because school-level factor scores were available only for the 2011 academic year, our analyses used data only in this year. Equation 1 describes our general model:

$$F_s = \beta_0 + \beta_1 \mathrm{X}_s + \varepsilon_s. \tag{1}$$

The factor score, F_s, for school s is modeled as a function of a vector of school characteristics, X_s, and an error term, ε_s. Summarized in Table 1, covariates include school level (e.g., elementary), average absences, average suspensions, enrollment, proportion of students in different racial/ethnic groups, proportion qualifying for free/reduced lunch, proportion ESE, and proportion gifted.

Do different kinds of teachers report different kinds of instructional collaboration? To properly account for the nested structure of our data, we use three-level, multilevel regression models with time at Level 1 (2 years of time-varying factor scores), teacher at Level 2, and school at Level 3. Equation 2 summarizes these models:

$$COLLAB_{yts} = \beta_0 + \beta_1 X_{yts} + \beta_2 Y_t + \beta_3 SchLevel_s + \phi_y + \nu_s + \mu_{ts} + e_{yts}$$
$$\text{where } \nu_s \sim N(0, \sigma^2{}_{int:school}); \mu_{ts} \sim N(0, \sigma^2{}_{int:teacher}). \tag{2}$$

In this specification, we model a teacher's collaboration factor score ($COLLAB_{yts}$) in year y, for teacher t, in school s, as a function of a fixed intercept, β_0, a vector of time-varying teacher and school characteristics, X_{yts}, a vector of time-invariant teacher characteristics, Y_t, school level (e.g., elementary, middle) $SchLevel_s$, a year fixed-effect, ϕ_y, and mutually independent random effects, associated with the schools, ν_s, teachers, μ_{ts}, and time, e_{yts}. Time-varying teacher characteristics include a dummy for level of experience (indicators for first 20 years; teachers with more than 20 years are the reference group) and job category (instructional, ESE, vocational, other); time-varying school characteristics include enrollment, average student absences, average student suspensions, and percentage of students from different racial/ethnic groups, who qualify for free/reduced lunch, who are ESE (nongifted), and who are gifted. Time-invariant teacher characteristics include gender, race/ethnicity, and highest degree (e.g., BA, MA). School covariates are summarized in Table 1; teacher covariates are summarized in Table 2.

Is the average quality of faculty collaboration associated with school achievement? To address this research question, we use OLS regression to estimate a school's value-added score in math or reading as a function of collaboration factor scores. Again, we run models only for the 2011 academic year. Equation 3 describes our general model:

$$Value\text{-}added_s = \beta_0 + \beta_1 F_s + \beta_2 \mathrm{X}_s + \varepsilon_s. \tag{3}$$

The value-added score for school s is modeled as function of that school's collaboration factor score, F_s, a vector of school characteristics, X_s (see Table 1 for covariates), and an error term, ε_s. We include school controls to account for characteristics of schools that might account for the observed relationships between collaboration and achievement. In alternative models, we include average teacher characteristics (aggregated at the school level) as controls. Due to space constraints, we do not report them here; results were similar and are available upon request.

Is a teacher's own collaboration quality or the average collaboration quality of her colleagues associated with her students' achievement? We conceptualize, and attempt to model, collaboration as simultaneously a collective and individualistic phenomenon. The quality of collaboration experienced by any individual is a product of both that individual's contributions to collaboration as well as the contributions of collaborators that surround her. Our analytic approach then uses a number of strategies to model the conjoint contribution of these different components of collaboration, while attempting to disentangle the effects of the former from the latter. Specifically, we (1) center each teacher's collaboration factor at the school-level mean, (2) use multilevel models to account for the nesting of individual collaboration within school collaboration, and in alternative models (3) control for any interaction effects between teacher- and school-level collaboration factors (described below).

To estimate a given teacher's value-added as a function of both her own and her school's collaboration factor scores, we use a three-level, multilevel model with time at Level 1 (2 years of value-added),[3] teacher at Level 2, and school at Level 3.[4] Teacher-level collaboration scores were school-mean centered; school-level factor scores were grand-mean centered. Equation 4 summarizes our general model specification:

$$VAM_{yts} = \beta_0 + \beta_1(COLLAB_{yts} - \overline{COLLAB}_{ys}) + \beta_2(COLLAB_{ys} - \overline{COLLAB}_{y}) + \beta_3 X_{yts} + \beta_4 Y_t + \beta_5 SchLevel_s + \phi_y + \nu_s + \mu_{ts} + e_{yts},$$
$$\text{where}\ , \nu_s \sim N(0, \sigma^2{}_{int:school}); \mu_{ts} \sim N(0, \sigma^2{}_{int:teacher}). \quad (4)$$

.

In this specification, we model the value-added to student achievement (VAM_{yts}) in math or reading in year y, for teacher t, in school s, as a function of a fixed intercept β_0, the teacher's own school-mean centered collaboration score ($COLLAB_{yts} - \overline{COLLAB}_{ys}$), the average of the collaboration scores of faculty in the teacher's school (grand-mean centered) ($COLLAB_{ys} - \overline{COLLAB}_{y}$), a vector of time-varying teacher characteristics, X_{yts}, a vector of time-invariant teacher characteristics, Y_t, school level (e.g., middle), $SchLevel_t$, a fixed effect for school year, ϕ_y, and mutually independent random effects associated with the schools, ν_s, teachers, μ_{ts}, and time, e_{yts}. See Table 2 for list of teacher covariates.

Because we assumed that a teacher's collaboration level would interact with her school's collaboration level, we initially included interaction terms between individual and school collaboration factors. However, a Deviance Likelihood Ratio Test indicated these additions did not improve our model fit. Moreover, none of the included interaction terms were statistically significant and estimates for main effects went virtually unchanged with their inclusion. We also initially included a random slope term for teacher-level collaboration factor scores because we assumed collaboration slopes would vary for individuals in different schools. However, Deviance Likelihood Ratio Tests indicated that random slope terms did not improve model fit.

Do teachers improve at greater rates in schools with better collaboration? To answer this question, we build upon an approach by Kraft and Papay (2014), which uses teacher fixed-effects regression models to estimate a teacher's value-added in math as a function of collaboration factors fully interacted with a quartic function for experience.[5] We prefer the teacher fixed-effects approach because it does a better job than other modeling approaches at ruling out alternative, noncausal explanations for observed relationships due to teacher sorting. In models without teacher fixed effects, for example, observed relationships between collaboration and achievement could be due to more promising teachers sorting into schools with better collaboration.[6] In contrast, using teacher fixed effects allows us to look within teachers, across schools to test whether returns to experience for the same teacher are greater when working in schools with better quality collaboration as compared to when working in schools with worse collaboration. Because we only had school-level factor scores for a single academic year (2011), we made the assumption that collaboration quality remained at the same levels for the prior and subsequent years. In support of this assumption, Kraft and Papay (2014) find school environmental measures, including collaboration, to be relatively stable across 6 years. Thus, we ran models for all teachers in all schools from 2009–2010 to 2011–2012. Equation 5 describes our modeling approach:

$$Value\text{-}added_{tys} = \beta_0 + \beta_1 F_s + \beta_3\big(f\big(EXP_{ty}\big)\big) + \beta_4\big(f\big(EXP_{ty}\big)\big) * F_s + \beta_3 \mathrm{X}_s + \beta_4 Y_{ty} + \phi_t + \nu_y + \varepsilon_{tys}, \text{ where}$$

$$f(EXP_{ty}) = EXP_{ty} + (EXP_{ty})^2 + (EXP_{ty})^3 + (EXP_{ty})^4. \qquad (5)$$

Value-added to achievement in either math or reading for teacher *t*, in year *y*, in school *s*, is modeled as a function of her school's collaboration factor score, F_s, a quartic function for her experience level, $f(EXP_{ty})$, an interaction between the collaboration factor and experience function, $(f(EXP_{ty}))*F_s$, a vector of school-level indicators (e.g., middle school), X_s, a vector of time varying and invariant teacher characteristics, Y_{ty}, teacher fixed effects, ϕ_t, year fixed effects, ν_y, and an error term, ε_{tys}. Teacher covariates are summarized in Table 2.

Results

What Kinds of Instructional Collaborations Exist in This Large Urban District?

Surveys revealed that the vast majority of teachers belonged to instructional teams and that they had favorable impressions about the quality of collaborations they experienced in these teams. Eighty-four percent of all respondents identified as being a part of "a team or group of colleagues that works together on instruction." The survey included a follow-up item that asked teachers generally about how helpful their instructional teams were for improving their instructional practice. Respondents' ratings were quite favorable. Almost 90% of respondents reported instructional teams as "helpful" (39%) to "very helpful" (49%); 10% reported them as "a little helpful" and only 2% as "not helpful."

For teachers who reported being members of instructional teams, the survey included a set of questions about (a) the extent to which different instructional domains were addressed in these teams and (b) the helpfulness of collaborations in these same domains. A summary of teachers' responses to these sets of questions is included in Table 4. We have ranked responses so that the areas of collaboration rated as most helpful and extensive are at top of the table.

Teachers indicated that collaboration was fairly extensive across instructional domains. On average, teachers reported that their teams covered all instructional domains "a little" to "in some depth." The extent of collaboration was greatest in two domains: reviewing formative assessments and developing instructional strategies. Reported levels were significantly greater in these domains than in others. The domains that were covered least extensively were "addressing classroom management/discipline issues" and "reviewing students' classroom work." The extent of collaboration in these areas was significantly less than in other domains.

In addition to being extensive, teachers felt that collaboration in instructional teams was also quite helpful. Teachers reported collaboration in all instructional domains to be, on average, "helpful" to "very helpful." However, they found collaboration about "developing instructional strategies" to be most helpful and significantly more helpful than collaboration in other domains. That this domain was also among the most extensively covered seems promising; it suggests that teams had attended to areas of teaching that members perceived to be useful. We considered whether teachers rated this domain as most helpful simply because they focused extensively on it. However, the forms of collaboration covered most extensively were not always perceived as most helpful. For example, "reviewing state test results" and "reviewing formative assessments" were covered extensively but were considered to be relatively less helpful.

The two domains that were covered least extensively, "addressing classroom management/discipline issues" and "reviewing students' classroom work," were also perceived to be the least helpful, and significantly less helpful than collaboration in the next highest domain. Given recent evidence for meaningful relationships between teachers' quality of classroom management and their effectiveness at raising student achievement (Grossman, Cohen, Ronfeldt, & Brown, 2014), we expected collaboration around classroom management to be viewed as more helpful. We also expected teachers to perceive collaboration around students' classroom work more favorably, given recent evidence of strong effects of professional development focused on student work (Gallimore et al., 2009; Saunders et al., 2009). Even so, these results are affirming in the sense that teams seem to be spending less time on "less helpful" domains and more time on "more helpful" domains, which suggests that they are able to make good use of the limited time they have to collaborate.

As described in the "Methods" section, we used the set of survey items described in Table 4 to construct the main predictors in our analyses, a set of factors signaling different kinds of instructional collaboration. Across factors, the majority of variation—between 84% and 87%—was within schools. Part of the explanation for why so much of the variation is within schools may be due to measurement error, which mostly gets attributed to within-school variation. Even so, these findings suggest substantial differences in collaboration may exist between individuals and/or between instructional teams within schools. In other words, between-school comparisons—the focus of most prior research—may mask important within-school differences.

Do Different Kinds of Schools Have Different Kinds of Instructional Collaboration?

Table 5 summarizes results from school-level OLS regression models estimating collaboration factors as a function of school characteristics. As indicated by the column titles, we used a different collaboration factor as the outcome measure for each of the models presented in this table. Results indicate that differences in collaboration exist between school levels. Compared to secondary (middle, high, 6–12) schools, teachers in elementary schools report, on average, better quality collaboration, in general, and about instructional strategies/curriculum and students, in particular. By contrast, the quality of collaboration about assessment was statistically similar between elementary and secondary schools.

Beyond school level, schools with higher proportions of students enrolled in ESE[7] (nongifted) programs tended to have lower factor scores, although estimates were statistically significant only on the "instruction/curriculum" factor. These results suggest that the quality of collaboration about instruction/curriculum was weaker on average in schools with more ESE students. Finally, teachers who worked in schools with larger enrollments reported better quality collaboration about instructional strategies/curriculum but worse quality collaboration about students. One possibility is that larger enrollments required faculty to invest more in dimensions of instruction, such as curriculum development, aimed at the collective, while smaller enrollments allowed teachers to collectively attend to the needs of particular students.

Do Different Kinds of Teachers Report Different Kinds of Instructional Collaboration?

Table 6 summarizes results from multilevel regression models estimating a teacher's instructional collaboration factor score as a function of her own (teacher) characteristics and the characteristics of her school. Each model has a different collaboration factor as the outcome variable. In this table, the reference group includes teachers who are Hispanic, whose highest degree is a BA, and who are identified by the district as having "instructional" positions. These models focus on the subsample of survey

Table 5
Do Different Kinds of Schools Have Different Kinds of Collaborations?

Variables	General Collaboration	Instruction Collaboration	Students Collaboration	Assessment Collaboration
Middle school	–0.6049**	–0.4572*	–0.6362**	0.0953
	(0.215)	(0.177)	(0.196)	(0.205)
PreK–8	0.0041	–0.1904	0.1695	0.0509
	(0.175)	(0.212)	(0.133)	(0.186)
Grades 6–12	–0.6232	–0.8747~	–1.1937*	1.1798~
	(0.594)	(0.514)	(0.547)	(0.641)
High school	–0.7861**	–0.8240**	–0.3049	–0.1740
	(0.255)	(0.270)	(0.249)	(0.264)
Proportion female	0.9663	0.9505	0.8751	–0.2625
	(0.631)	(0.626)	(0.651)	(0.638)
Proportion free/ reduced lunch	0.6210	0.0116	0.5069	0.6166
	(0.645)	(0.682)	(0.513)	(0.611)
Proportion White	–1.0769	–0.4881	–0.0052	–1.4678
	(1.009)	(1.081)	(1.043)	(0.990)
Proportion Black	0.3141	0.1025	0.4733	–0.0443
	(0.350)	(0.306)	(0.294)	(0.380)
Proportion Asian	–1.0668	4.1601	–0.0831	–6.9785
	(5.315)	(6.317)	(5.228)	(5.931)
Proportion Native American	42.8492	15.9668	–2.0735	65.0874
	(57.738)	(46.425)	(41.476)	(47.863)
Proportion multirace	0.1494	3.0972~	–4.3868***	1.3833
	(1.441)	(1.684)	(1.238)	(1.691)
Proportion LEP	–0.1716	0.1096	–0.0815	–0.3713
	(0.469)	(0.433)	(0.412)	(0.511)
Proportion ESE	–0.9808	–1.3511*	0.1544	–0.4169
	(0.832)	(0.637)	(0.530)	(0.738)
Proportion gifted, non-ESE	0.6795	–0.3550	1.6039	–0.0386
	(1.124)	(1.119)	(1.043)	(1.275)
Average absences	0.0183	0.0047	0.0436~	–0.0187
	(0.031)	(0.026)	(0.023)	(0.028)
Average suspensions	0.0236	0.0253	0.0315	–0.0201
	(0.084)	(0.067)	(0.081)	(0.073)
Enrollment per 100	–0.0000	0.0004**	–0.0003**	–0.0002
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	–0.7657	–0.5662	–1.0021	0.3225
	(0.552)	(0.665)	(0.618)	(0.660)
N	336	336	336	336
R-squared	0.2097	0.1117	0.2524	0.1618

Note. Standard errors are in parentheses. Elementary schools and Hispanic students were reference groups.
~$p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$.

respondents who also had value-added. To investigate observed differences further, we ran alternative models by changing the reference groups and using the full sample of survey respondents. Throughout this section, we describe findings from alternative models but, due to space limitations, do not report point estimates.

The results indicate that different kinds of teachers indeed experience different kinds of instructional collaboration. Female teachers report 27% of a standard deviation better quality collaboration in general. Looking at collaboration in different instructional domains, they report significantly better collaboration about instructional strategies and curriculum (by 25% of a standard deviation) and about assessment (by 16% of a standard deviation). The perceived quality of collaboration about students, however, is similar between males and females.

Findings also suggest differences in collaboration between individuals from different teacher race subgroups. In general, White teachers report lower levels of collaboration quality than Hispanic teachers, who in turn report lower levels than Black teachers. Across outcome measures, White teachers report worse collaboration than Hispanic teachers (reference group).[8] In fact, when compared to all racial subgroups (combined), White teachers had significantly lower scores on all collaboration measures except on the "instructional strategies and curriculum" factor, for which all teacher race subgroups reported statistically similar levels of collaboration quality. Differences were often quite substantial between racial subgroups, ranging from 17% ("students" factor) to 22% ("general" factor) of a standard deviation. With the exception of the "instructional strategies and curriculum" factor, Black teachers reported the highest levels of collaboration. Compared to Hispanic teachers, Black teachers had 15% of a standard deviation greater "general" and "assessment" factor scores and 17% of a standard deviation greater "students" factor scores; differences with White teachers were even greater.

Teachers with different highest degree levels also reported different levels of collaboration. In particular, teachers who had a BA as their highest degree reported stronger collaboration than teachers with other kinds of highest degrees. When compared to teachers with all other highest degree types (combined), teachers with a BA had significantly higher scores on the "general" factor (by 17% of a standard deviation), "instructional strategies/curriculum" factor (by 15% of a standard deviation), and "students" factor (by 13% of a standard deviation); degree level groups reported statistically similar levels of collaboration about assessment.

Finally, we find teaching experience to be unrelated to most collaboration factors. Because our continuous measure for teaching experience predicted collaboration about assessment at the $p < .1$ level, we decided to investigate differences between more and less experienced teachers further. In particular, we compared teachers with 15 or more years of experience to less

Table 6
Do Different Kinds of Teachers Report Different Kinds of Collaborations?

Variables	General Collaboration	Instruction Collaboration	Students Collaboration	Assessment Collaboration
Female	0.2650***	0.2570**	0.04560	0.1617*
	(0.07995)	(0.08471)	(0.08337)	(0.07179)
White	–0.1401~	–0.05139	–0.07446	–0.1307~
	(0.07678)	(0.08116)	(0.07999)	(0.06877)
Black	0.1477~	–0.04236	0.1740*	0.1494*
	(0.07918)	(0.08376)	(0.08246)	(0.07103)
Asian	–0.1357	–0.04589	–0.2621	0.09552
	(0.2485)	(0.2632)	(0.2595)	(0.2226)
Native American	0.6718	0.4095	0.6633	0.05847
	(0.6655)	(0.7130)	(0.6979)	(0.6036)
AA degree	–0.9979	–1.7459~	0.1235	0.04293
	(0.9409)	(1.0114)	(0.9870)	(0.8580)
PhD degree	–0.3449	–0.2395	–0.4223~	0.1257
	(0.2189)	(0.2324)	(0.2288)	(0.1966)
MA degree	–0.1419*	–0.1349*	–0.1023	–0.006433
	(0.06162)	(0.06638)	(0.06503)	(0.05608)
Special degree	–0.2693**	–0.1157	–0.2603*	–0.1105
	(0.1024)	(0.1101)	(0.1080)	(0.09297)
NA degree	–0.5905~	–0.3819	–0.3568	–0.1903
	(0.3328)	(0.3761)	(0.3634)	(0.3149)
Job: ESE	0.07829	0.02930	0.1277	–0.04250
	(0.08345)	(0.08895)	(0.08753)	(0.07519)
Job: vocational	0.6408~	0.5270	0.6452~	–0.2780
	(0.3422)	(0.3791)	(0.3691)	(0.3179)
Job: other	–0.05065	0.04399	–0.1434	–0.02331
	(0.1953)	(0.2108)	(0.2059)	(0.1782)
Experience	–0.0001915	–0.001341	–0.002547	0.004490~
	(0.002969)	(0.003146)	(0.003097)	(0.002666)
N	1,122	1,122	1,122	1,122
School controls	X	X	X	X

Note. Standard errors are in parentheses. These models are for only the subsample of teachers with value-added information. Estimates are from multilevel regression models with time at Level 1, teacher at Level 2, and school at Level 3. Hispanic teachers are the reference group for race/ethnicity; teachers with a BA degree are the reference group for highest degree earned; teachers who were assigned as having "instructional" jobs by the district are the reference group for job categories; "Job: ESE" refers to educators of students receiving ESE services.
$\sim p < .1$. $*p < .05$. $**p < .01$. $***p < .001$.

experienced teachers, finding the former group to report, on average, 12% of a standard deviation higher level of collaboration about assessment. We were somewhat surprised by these findings, as we hypothesized less experienced teachers would likely engage in more extensive collaboration and likely view collaboration as more helpful. Consistent with our hypothesis, in separate analyses using the full sample of survey respondents (relaxing the constraint of also having value-added scores), teachers with 15 or more years of experience reported significantly *lower* quality collaboration about instructional strategies and curriculum (by 16% of a standard deviation).

One possible explanation for observed differences in collaboration between kinds of teachers is that certain kinds of teachers sort into schools with certain kinds of collaboration. For example, female teachers may not actually be better collaborators but just happen to work in schools with better collaboration. To investigate this possibility, we tested whether coefficients on teacher-level collaboration decreased substantially after including school-level collaboration measures. Estimates remained similar, however, suggesting the observed relationships are not due to sorting of certain kinds of teachers into more collaborative settings.

Though our results indicate significant differences in teacher-level collaboration scores between different kinds of teachers, a limitation of using self-reported survey data to signal collaboration quality is that we cannot distinguish differences in *perceptions* of quality from *actual* quality differences. We cannot know, for example, whether male teachers are actually engaging in less extensive and less helpful collaboration or just perceiving this to be the case. It is possible that male and female teachers are engaging in qualitatively similar kinds of collaboration but that their subjective experiences differ. Still, there are reasons that the subjective experience of collaboration quality is important to study. For instance, a teacher's subjective experience of collaboration quality likely will influence future decisions to engage in collaboration, whether or not a more objective measure provides a similar assessment of quality. Another possible limitation of using self-report data is that, even if different subgroups of teachers have similar subjective experiences of collaboration quality, they might interpret, and thus respond to, the survey scales differently.

Is the Average Quality of Faculty Collaboration Associated With School Achievement?

Table 7 summarizes results from regression models estimating school-level value-added in math (left side) and in reading (right side) as a function of collaboration factors. In Model 1, we estimate value-added as a function of our domain-general factor score; in Model 2, we enter domain-specific collaboration factor scores together.

Results in Table 7 indicate that all collaboration factors positively and significantly predict math value-added. In Model 1, schools with 1 standard deviation better quality of "general" collaboration had 43% of a standard deviation

Table 7
Estimating School-Level Value-Added in Math and Reading as a Function of Different Kinds of Instructional Collaboration (2011 Only)

	Math Value-Added		Reading Value-Added	
Variables	Model 1	Model 2	Model 1	Model 2
General collaboration	0.4254***		0.1789***	
	(0.081)		(0.050)	
Instruction collaboration		0.2971**		0.1266*
		(0.099)		(0.051)
Students collaboration		0.1763*		0.1091*
		(0.079)		(0.050)
Assessment collaboration		0.2514***		0.0749
		(0.069)		(0.058)
N	335	335	336	336
R squared	0.1328	0.1377	0.0554	0.0562

Note. Robust standard errors are clustered at the school level (in parentheses); school response rates on surveys are used as probability weights.
*$p < .05$. **$p < .01$. ***$p < .001$.

higher math value-added. In Model 2, schools with 1 standard deviation better quality collaboration about instruction, assessment, and students had, respectively, 30%, 25%, and 18% of a standard deviation higher math value-added.

Table 7 indicates that the quality of collaboration in instructional teams also predicts reading value-added. In Models 1 and 2, all collaboration factors are positively associated with reading value-added and all but the "assessment" factor are significant predictors. As is often the case with value-added measures, point estimate magnitudes in reading were smaller than those in math. In Models 1 and 2, a 1 standard deviation increase on collaboration measures is associated with between 8% and 18% of a standard deviation increase in reading value-added.

Results from this section indicate that schools that have instructional teams engaged in better collaboration also have higher achievement gains in both math and reading. This evidence is suggestive that instructional collaborations have positive effects on students' achievement gains; the design of the study, however, does not lend itself to drawing causal conclusions. As we investigate below, a number of alternative, noncausal explanations are possible.

Testing validity threats. Although we cannot rule out all alternative explanations, we tested three threats to a causal interpretation of the impact of collaboration on student achievement gains. Our motivation for these tests was largely a concern over omitted variable bias: That something correlated with collaboration might explain observed relationships between collaboration and value-added. Results from these tests, described below, bolster the case for drawing causal inferences.

First, we tried adding aggregate (school-level) teacher characteristics as additional school controls in our models. Our purpose for adding these controls is that other unobserved school characteristics, like average teacher experience, could still account for observed relationships. Point estimates remained similar and, in some cases, were even greater, suggesting average teacher characteristics did not explain the observed relationships.

Second, we included the set of controls for school characteristics listed in Table 1. Our rationale for doing so was that teacher collaboration could vary with other school characteristics that explain the differences in achievement between schools. The inclusion of these controls reduced point estimates on collaboration measures somewhat, suggesting that school characteristics explain some of the observed relationships. Even with these controls, however, point estimates remained generally large in magnitude and statistically significant, suggesting school characteristics do not explain observed relationships between collaboration and achievement. Even so, as we consider next, other unobserved characteristics, like school leadership, might still explain them.

Third, we constructed, and controlled for, a number of measures based on survey items about teachers' satisfaction with dimensions of their work beyond collaboration in instructional teams. These included measures for how well respected teachers felt by leaders in their schools as well as a general measure of loyalty to and satisfaction with one's school. We included these controls to account for the possibility that respondents who were more satisfied with their schools for reasons apart from collaboration might report collaboration as being more helpful. Additionally, by controlling for teachers' satisfaction with other, noncollaboration dimensions of their work, we improved estimates for collaboration-specific effects. Point estimates were similar with the inclusion of these controls and, in some cases, even larger in magnitude.

Is a Teacher's Own Collaboration Quality or the Average Collaboration Quality of Her Colleagues Associated With Her Students' Achievement?

Assuming the observed school-level relationships between collaboration and value-added are causal, by what mechanism might collaboration impact student achievement? A likely explanation is that collaboration improves the instruction of individual teachers who participate in these collaborations; better instruction, in turn, improves students' achievement. If so, we would expect individuals who report engaging in more helpful and extensive collaborations to benefit most. Table 8 summarizes results from multilevel models estimating a teacher's value-added in mathematics (left side) and reading (right side) as a function of that teacher's collaboration factor scores and her school's aggregate collaboration factor scores. More specifically, we use a three-level model with time at Level 1 (2 years of value-added), teacher at Level 2, and school at Level 3. This approach allows us to estimate the effects of teacher-level collaboration simultaneous with school-level collaboration, while accounting for the nesting of the former within the latter. Teacher-level collaboration measures are school-mean centered; school-level measures are grand-mean centered.

Math value-added. Beginning with teacher-level collaboration factors, Table 8 results suggest that individual teachers who reported engaging in better quality collaboration in general, across instructional domains, had higher math value-added. The point estimate indicates that engaging in 1 standard deviation better general collaboration is associated with 9% of a standard deviation higher math value-added. Although the coefficient is only marginally statistically significant ($p < .1$), the difference in performance is meaningful. Based upon findings from Boyd et al. (2006, 2008), these estimates are roughly comparable to half a year of initial teaching experience. When we looked at collaboration in different instructional domains, only collaboration about assessment predicted math value-added at significant and meaningful levels. The results indicate that an increase of 1 standard deviation in quality of collaboration about assessment is associated with an increase of 15% of a standard deviation in math value-added, comparable to typical gains from an initial year of teaching experience. These results suggest that teachers who reported engaging in better quality collaboration demonstrated better achievement gains than peers who engaged in worse quality collaboration, even after controlling for the effects of school-level collaboration.

Moving to estimates on school-level collaboration factors, Table 8 indicates that both kinds of collaboration that were significant at the teacher level (general and assessment) were also significant at the school level. Teachers who worked in schools with 1 standard deviation higher scores for collaboration in general and for collaboration about assessment had, respectively, 15% and 14% of a standard deviation higher math value-added. Teachers have higher math growth not only when they themselves engaged in better general collaboration and better collaboration focused on assessment but also, as these results suggest, when they worked in schools that had better collaboration in these areas. Said another way, even after controlling for a teacher's own level of collaboration, teachers performed better when in schools that had better collaboration in these areas.

Additionally, we find that teachers who worked in schools with better quality collaboration about students tended to be more effective at raising student achievement. However, a teacher's own reported level of collaboration about students was unrelated to her value-added. In other words, when

Table 8
Estimating Teacher-Level Value-Added as a Function of Instructional Collaboration

	Math Value-Added		Reading Value-Added	
Variable	Model 1	Model 2	Model 1	Model 2
General collaboration (teacher level)	0.09059~ (0.04883)		0.008629 (0.04029)	
General collaboration (school level)	0.1463* (0.06987)		0.1112~ (0.06019)	
Instruction collaboration (teacher level)		0.01474 (0.04442)		0.08079* (0.03710)
Instruction collaboration (school level)		–0.01443 (0.06389)		0.07739 (0.05449)
Students collaboration (teacher level)		0.03968 (0.04579)		–0.04456 (0.03836)
Students collaboration (school level)		0.1906** (0.07039)		0.02613 (0.05861)
Assessment collaboration (teacher level)		0.1446** (0.05147)		–0.04304 (0.04377)
Assessment collaboration (school level)		0.1439* (0.06835)		0.07350 (0.05517)
N	542	542	674	674
Teacher characteristics	X	X	X	X
School controls	X	X	X	X

Note. Teacher-level instructional collaboration factors are school-mean centered; school-level instructional collaboration factors are grand-mean centered.
~$p < .1$. *$p < .05$. **$p < .01$.

teachers themselves engaged in better quality collaboration, they were no more or less effective than peers who engaged in worse quality collaboration; however, when individuals taught in schools with better collaboration among colleagues, they were more effective than teachers who worked in schools with worse quality collaboration in this area, even after accounting for a teacher's individual collaboration level. Said simply, the collective contribution to collaboration about students was positively related to the achievement gains of a given teacher while her individual contribution to collaboration was unrelated. How can this be? One possibility is that schools with a culture of collaboration about particular students have better mechanisms for triaging student issues and distributing teaching responsibility across the community in ways that help students to achieve. In this way, a given teacher's performance can benefit from the quality of collaboration among colleagues regardless of her own contribution in this regard. By contrast, high levels of independent contributions to collaboration about students, apart from school-wide contributions in this area, are unrelated to achievement gains. Regardless, finding the level of collaboration among colleagues to be more predictive of an individual's value-added than her own level of contribution raises some questions about the appropriateness of interpreting value-added scores as a teacher "effect."

Reading value-added. As illustrated on the right side of Table 8, few coefficients on instructional collaboration factors, either at the teacher or school levels, were statistically significant predictors of reading value-added. Among teacher-level factors, only the measure for collaboration about instructional strategies and curriculum was a significant predictor. Teachers who engaged in 1 standard deviation higher levels of collaboration about instructional strategies and curriculum were 8% of standard deviation more effective at raising student achievement in reading, an amount of growth in performance that is comparable to one half of a year of initial teaching experience. Among school-level factors, none were significant at the $p < .05$ level, and only the general collaboration factor was significant at the $p < .10$ level. Although not statistically significant, the "general," "instructional strategies/curriculum," and "assessment" school-level factors all had point estimates of 0.07 or greater, amounts that prior research suggests to be meaningful—comparable to one half of a year of initial teaching experience.

In alternative models, we included interaction terms between teacher- and school-level collaboration measures because we wanted to examine whether the effects of teacher-level collaboration varied by the effects of school-level collaboration and vice versa. The inclusion of interaction terms did not improve our model fit and did not significantly alter main effects; moreover, none of the interaction terms were statistically significant. As a result, this section focuses on results from models without interaction terms. Even so, the fact that interaction terms were not statistically significant is worth highlighting. These null results suggest that the positive relationships between school-level collaboration and value-added are not statistically different for more and less collaborative individuals. Additionally, these results indicate that, even when working in schools with better or worse quality collaboration, the benefits of being a more collaborative individual persist.

Do Teachers Improve at Greater Rates in Schools With Better Collaboration?

If collaboration quality indeed causes student achievement to improve, it would likely do so first by improving the effectiveness of individual teachers engaged in collaboration. To investigate, we next test whether teachers' rates of improvement are greater in schools with better quality collaboration. We use teacher fixed-effects regression models to estimate a teacher's value-added as a function of collaboration factors fully interacted with a quartic function for experience. In Table 9, we report coefficients for the

Table 9
Estimating Teacher Value-Added Differential Returns to Experience With Increasing Collaboration Quality in Schools: Teacher Fixed Effects Models 2010–2012

	Math Value-Added		Reading Value-Added	
Variable	Model 1	Model 2	Model 1	Model 2
General Collaboration	0.06648*		0.04756~	
× Experience	(0.02690)		(0.02522)	
Instruction Collaboration		0.01905		0.01104
× Experience		(0.02634)		(0.02389)
Students Collaboration		0.03606		0.04471
× Experience		(0.02602)		(0.02724)
Assessment Collaboration		0.06176*		0.02878
× Experience		(0.02549)		(0.02414)
N	6,682	6,682	7,880	7,880
Teacher fixed effects	X	X	X	X
School controls	X	X	X	X
Year indicators	X	X	X	X

Note. Standard errors are in parentheses. These models include teachers from the 2010–2012 academic years. Experience is entered as a quartic function, fully interacted with our measures for instructional collaboration; we present estimates only on the linear component interacted with collaboration. Models use the same 2011 school collaboration factors in 2010 and 2012, making the assumption that schools with better collaboration in 2011 also have better collaboration in the subsequent and prior years.
~$p < .1$. *$p < .05$.

interactions between collaboration factors and the linear term of the quartic experience function.

Coefficients on all interaction terms are positive, indicating that a given teacher's rate of improvement increases when she works in schools with better collaboration as compared to when she works in schools with worse collaboration. Regarding math value-added (left side), returns to experience are significantly greater in schools with better quality of collaboration, in general, as well as in schools with better collaboration about assessment, in particular. Estimates suggest that a 1 standard deviation increase in the quality of these forms of collaboration predicts a 6% to 7% of a standard deviation increase in returns to experience. In other words, when a teacher moves to a new school that has better collaboration by 1 standard deviation over her prior school, she gains an additional 6% to 7% of a standard deviation increase in math value-added each year, comparable to average gains from one half a year of initial teaching experience.

Reading value-added coefficients also trended positive, but none of the point estimates were statistically significant at the $p < .05$ level. The coefficient on the "general" factor was significant at the $p < .1$ level, and all estimates, except for collaboration about instruction, were similar and nontrivial in magnitude—3% to 5% of a standard deviation increase in returns to experience. These results may provide weak evidence for better returns to experience when teachers work in schools with better quality collaboration. For both math and reading value-added, the differential returns to experience were smallest with increases in the quality of collaboration about instructional strategies and curriculum. This is notable given that teachers typically rated collaboration in this domain as more helpful than other domains (see Table 4).

Discussion

As the push to create communities of practice in schools takes hold, understanding the increasingly collaborative nature of teachers' work takes on heightened importance. This study describes that work and investigates how it varies by school context and teacher characteristics. It also sheds light on how teacher collaboration contributes to teacher improvement and student achievement. Looking across analyses, results suggest that collaboration in instructional teams is associated with gains on both fronts. Schools and teachers that have better quality collaboration across instructional domains (i.e., "general" collaboration factor) also have higher achievement gains, and usually at statistically significant and meaningful levels.

Does collaboration about some instructional domains affect student achievement more than collaboration about other domains? The answer to this question is less clear and seems to vary somewhat by subject area (reading or math) and by modeling approach. Coefficients on all forms of collaboration trended positive across models and tables, suggesting positive effects of collaboration regardless of its instructional focus. Even so, collaboration about assessment was most often significantly predictive of achievement gains across math and reading. In reading, collaboration about instructional strategies and curriculum also predicted achievement gains.

Given that our measures for school and teacher performance are based upon student achievement gains, perhaps it is unsurprising that collaboration about student assessment is more often predictive than collaboration about other instructional domains. That is, one might expect the quality of collaboration about a specific phenomenon to improve performance related to that phenomenon. In fact, finding collaboration around assessment to be more often predictive of achievement gains than collaboration in other domains may strengthen the case that observed relationships are indeed causal.

Despite providing suggestive evidence that collaboration in instructional teams can improve teacher and school performance, the design of our study does not permit us to draw causal conclusions. First, observed relationships between teacher collaboration and student achievement could be explained by differences in teacher or school characteristics. By controlling for a comprehensive school and teacher characteristics and by conducting tests of various threats to causal interpretations, we have tried to rule out alternative explanations for observed relationships. Although we find significant relationships to persist across analyses, it is still possible that other unobserved teacher and school characteristics could explain this consistent pattern of associations.

Second, our correlational evidence does not permit us to rule out reverse causality; that is, strong achievement gains may actually cause teachers to collaborate better. Our teacher fixed-effects analyses present some evidence to counter reverse causality. Namely, these analyses indicate that the same teacher improved at greater rates when she worked in schools with better collaboration than when she worked in schools with worse collaboration. It is difficult to imagine a scenario in which a transferring teacher's increased rate of performance in her new setting could cause average improvements in collaboration quality among colleagues in that setting. Even so, we cannot rule out such an explanation, especially given limitations in our modeling approach; these models assumed average school-level collaboration quality measured in 2011 would be constant from 2011 to 2012. Future studies should gather longitudinal information to test whether increases in collaboration temporally precede increases in performance.

Third, using self-reported measures for instructional quality presents further challenges to a causal interpretation. It is possible, for example, that high-performing teachers are generally more psychologically positive about their work and their workplace. If so, then they may rate the same instance of collaboration more positively than a colleague who is performing worse. By controlling for measures of teachers' satisfaction with aspects of work likely unrelated to collaboration, we made progress in accounting for this alternative explanation. Still, future research might gather large-scale observational data on the collaborative behaviors, discourse, norms, and structures of instructional teams to test whether specific, observable features of collaboration are related to achievement gains. Such a study could build upon prior case study research documenting the routines and norms that characterize promising PLCs (Horn & Little, 2009; Little, 2003).

Despite these aforementioned limitations, three findings from this study bolster the case for making causal inferences. First, individuals who reported engaging in better quality collaboration demonstrated better achievement gains than peers who engaged in worse quality collaboration, even after controlling for the effects of school-level collaboration. If the effect of collaboration quality is indeed causal, then one would expect individuals who engage in better quality collaboration to benefit more. Second, as mentioned above, collaboration about assessment, the instructional domain most directly related to the outcome being measured, was more often predictive of test score gains than collaboration in other instructional areas. If collaboration is causing achievement gains, then we would expect collaboration related to this outcome to have the strongest impact on it. Third, the returns to a single year of teaching experience were significantly greater for a teacher working in a school with better collaboration as compared to the same teacher working in a school with worse collaboration. If collaboration quality has a causal impact on teachers' performance, then one would expect the same teacher to improve more when she works in schools with better collaboration than in schools with worse collaboration. Moreover, these results counter one of the most likely threats to a causal explanation: that more promising teachers sort into schools with better collaboration. Because these models included teacher fixed effects, comparisons are within teacher, across time.

Implications

Assuming our findings indeed reflect causal effects of collaboration on teachers' and schools' effectiveness at improving student achievement, they suggest that a promising approach to educational improvement is through increasing the quality of collaboration that occurs in instructional teams. Thus, our findings offer empirical support for the continued use of instructional teams, already common in the district, and for ensuring that these teams have opportunities to engage extensively in the kinds of collaboration that they find to be helpful.

Ours is not the first study to suggest that collaboration among instructional teams is a promising policy focus. In their quasi-experimental intervention study, for example, Gallimore et al. (2009) offered large-scale evidence that promoting quality collaboration in grade-level instructional teams can increase student achievement. Although this study provided strong evidence that carefully designed, well-resourced, and ongoing interventions for promoting collaboration can improve student achievement, it did not address whether more typical collaboration among teachers is also beneficial. Ours is the first large-scale study of which we are aware to examine the kinds of collaborations that naturally occur in instructional teams and their effects on school and teacher performance. Results suggest that, indeed, naturally occurring collaboration—taking various forms and serving various functions within and across schools—also promotes student achievement.

Although our results suggest that collaboration of all stripes has an average positive impact on achievement, we wondered whether effects depend upon collaboration content. Our "general" factor, however, signaling collaboration that spans a range of instructional domains, was most consistently predictive of student achievement gains across models and subject areas. These estimates likely were bolstered by the greater reliability of the General scale relative to the other factors. Nonetheless, a potential implication is that promoting high-quality collaboration across a range of instructional domains is a better way to improve student achievement than promoting collaboration in any single domain.

While collaboration across a range of instructional domains is likely beneficial, time and resources for collaboration are limited, so teachers and schools often must prioritize some instructional topics over others. To this point, we find that collaboration about assessment was more often predictive of school and teacher performance than collaboration about other instructional domains. When the goal is to raise student test scores, our findings suggest that districts and schools benefit most by structuring collaboration in instructional teams around local formative assessments and state assessments. Does this mean that building collaboration around assessments is good policy? Not necessarily. It is possible, for example, that test score gains may have resulted from an excessive focus on test preparation, possibly at the expense of focusing on other educationally meaningful topics. Finding collaboration about assessments to predict better performance on assessments is not so surprising. Had our dependent variable been a different educational outcome, for example, teachers' pedagogy or students' critical thinking, it is possible that collaboration about assessment would not have been as predictive. Future research should continue to investigate whether different educational outcomes are more responsive to collaboration with different foci.

This study also has implications for understanding the mechanisms by which collaboration influences student achievement. We investigate two explanatory mechanisms by which student achievement may benefit from collaboration among teachers: an "individualistic" and a "collectivist" mechanism. Specifically, we examine the quality of collaboration experienced by individual teachers apart from, and in conjunction with, collective (school-level) collaboration quality, thus allowing us to test both explanatory mechanisms (see Table 8).

One possible way that collaboration impacts achievement is that each individual teacher gains useful knowledge as she collaborates with colleagues about instruction, which in turn improves her individual performance. According to this "individualistic" mechanism, teachers who make the best use of available collaboration will likely benefit most; students of teachers who take no part in collaboration will show no gains in achievement. Consistent with the above explanation, we find some evidence that individual teachers have better achievement gains when they engage in better quality collaboration than their peers, even after controlling for the effects of collective levels of collaboration in their school settings. In other words, an individual teacher has better achievement gains as a result of engaging in high levels of collaboration, whether or not her colleagues' average level of collaboration is also high. Finding nonsignificant interactions between teacher- and school-level collaboration terms offers further support for this explanation. The implication is that individuals benefit from seeking out and taking advantage of available collaborative resources, regardless of how rich these resources may be. Even in the absence of school-wide collaborative structures, then, individual efforts to collaborate should not go overlooked; school leaders and teachers should find ways to encourage such efforts.

Another possibility, though, is that the students of a given teacher might experience greater achievement gains because of the surrounding collective levels of collaboration, regardless of the individual contribution of the teacher in question. We call this the "collectivist" mechanism. An extreme example might help to illustrate this distinction. One can imagine a scenario in which, as a collective, a school faculty engages in high-quality collaboration to which a given teacher makes minimal contribution. According to a collectivist explanatory mechanism, the students of the nonparticipating teacher would still benefit from the collective engagement of her colleagues. For example, these students may learn study skills from teacher colleagues that spill over to the classroom of the nonparticipating teacher. In contrast, these students would show no improvement according to the individualist framework described above.

Our findings provide evidence in support of the collectivist mechanism as well. Even after controlling for an individual's own level of collaborative engagement, we find collective levels of collaboration to predict her own students' performance. In fact, in Table 8, coefficients on collective (school-level) collaboration terms tend to be larger in magnitude and more often statistically significant than coefficients on individual (teacher-level) collaboration terms. In the case of collaboration about students' needs, only collective (school-level) collaboration levels predicted a teacher's achievement gains; her own (teacher-level) engagement in collaboration about specific students was unrelated to the achievement gains of her students. Finding evidence for the collectivist mechanism suggests that school leaders and practitioners work to establish school-wide structures and systems for promoting teacher collaboration.

These findings also suggest the possibility that the achievement gains of a given teacher are, at least in part, a product of the collaborative environment that surrounds her, regardless of the relative extent to which that teacher engages with said environment. This possibility raises some concern over the appropriateness of attributing achievement gains, in a given teacher's classroom, to her own performance or effect. This concern is consistent with results from Jackson and Breugmann (2009), who estimate about 20% of a given teacher's value-added is explained by the value-added of her immediate grade-level team colleagues. These authors postulate that collaboration with colleagues likely explains the spillover effect they observed, though they did not observe teachers' collaboration. Our findings extend prior work by offering direct evidence that collaboration is one mechanism by which colleagues can contribute to teachers' performance or, more precisely, the performance of their students. We make this distinction, in part, to punctuate the possibility that these results may complicate the attribution of value-added scores to effects of individual teachers.

We began this section with the assumption that the observed relationships between collaboration quality and student achievement are causal. Before taking seriously the implications discussed above, more research is needed to further interrogate whether this is indeed the case. Assuming it is, future research should also investigate the mechanisms by which collaboration impacts student achievement. One possibility is to collect observational data on teachers' instruction to test whether instructional quality mediates the relationship between teacher collaboration and student achievement. Finally, it would be helpful to know whether certain observable features of collaboration improve the quality of collaboration experienced by teachers or student achievement.

Another potential concern is that the district context we have studied is unique, preventing the results from generalizing elsewhere. In response to local or state policies, for example, schools might support higher quality teacher collaboration in MDCPS than occurs in other districts. We have, however, a number of reasons to suspect our results may be relevant to other districts. First, national evidence suggests that teachers across the United States engage in levels of collaboration, in general, and in instructional teams, in particular, that are similar to what we observe in MDCPS (MetLife Foundation, 2009). Second, inquiries to district personnel revealed no district policies during the period of the study that would have made collaboration or its relationship with achievement unique to MDCPS. The district supports the establishment of PLCs within schools, and the district's lowest performing schools (called "transformational" schools) in particular utilize literacy and other instructional design teams as an improvement strategy, but these approaches are far from specific to the district. One potentially unique circumstance was the district's implementation of a new induction plan in 2010–2011, which primarily affected teachers in their first 3 years of experience. Among many changes, the plan required that early career teachers participate monthly in PLCs. However, similar kinds of induction programs have become common across the United States (Goldrick, Osta, Barlin, & Burn, 2009). Moreover, our results were not driven by early career teachers; in fact, we find that experienced teachers sometimes report more extensive and helpful collaboration than early career teachers. Nevertheless, we encourage future research to investigate whether the relationships between teacher collaboration and student achievement we observed also exist in other district contexts.

In conclusion, Lortie (1975) observed long ago that a persistent and pernicious challenge to the profession is the isolation of its teachers who do not commonly engage in teaching as collective work. In response to this challenge, many practitioners and policymakers have since worked to promote collaboration in schools, including the creation of PLCs. Our study suggests these efforts may be paying off. In the district we studied, about 85% of teachers now identify as being a part of "a team or group of colleagues that works together on instruction" and report that collaboration in these teams is quite extensive and helpful. In addition to being good for teachers, our results also suggest that these instructional collaborations benefit students. Student achievement gains are greater in schools with stronger collaborative environments and in classrooms of teachers who are stronger collaborators.

Notes

This research was supported in part by a grant from the Institute of Education Sciences at the U.S. Department of Education (R305A100286). We thank the leadership of the Miami-Dade County Public Schools for the assistance they provided in collecting the data used in this study. We are particularly indebted to Gisela Feild and her staff. We thank Mari Muraki for excellent data management. We also thank Susanna Loeb, John Papay, Matt Kraft, and our blind reviewers for helpful comments on earlier drafts of this article. All errors are the responsibility of the authors.

[1]This subset of teachers includes only about one-eighth of the larger sample because (a) there were low survey response rates, (b) the 2012 survey was administered only to a random sample of teachers in each school, and (c) there was loss of teachers due to missing data for any survey items included in factors.

[2]The eigenvalue measures the variance of all of the variables accounted for by a factor. A common criterion, the Kaiser criterion, is to drop all components with eigenvalues less than 1.

[3]About 20% of teachers in our sample are included in both years of our analysis. The remaining teachers had only had a single year of data, typically because they responded to only a single year of the survey or because their value-added scores were available in only 1 year.

[4]Intraclass correlations indicated that clustering at the teacher (ICC = 0.46) and school (ICC = 0.05) levels to be nonzero. These results justify our use of three-level models.

[5]Papay and Kraft (2011) demonstrate that a quartic function is appropriate for modeling the relationships between teachers' value-added and years of experience.

[6]We ran multilevel models with teachers at Level 1 and schools at Level 2, relaxing the restriction that teacher effects be fixed. Results were similar and are available upon request.

[7]The ESE classification covers a broad range of student populations, including students with emotional/behavioral disabilities, autism spectrum disorders, intellectual disabilities, specific learning disabilities, and visual/hearing impairments. Though gifted students are included in this district classification, we analyze "proportion gifted" separately.

[8]In separate models, using the full survey sample (not restricted to having value-added), coefficients were significant ($p < .05$) on all factors except "collaboration about instruction."

References

Bryk, A., Camburn, E., & Louis, K. S. (1999). Professional community in Chicago elementary schools: Facilitating factors and organizational consequences. *Educational Administration Quarterly*, *35*(5), 751–781.

Carroll, T. G. (2007). Teaching for the future. In B. Wehling & C. Schneider (Eds.), *Building a 21st century U.S. education system* (pp. 46–58). Washington, DC: National Commission on Teaching and America's Future.

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research & Perspective*, *9*(4), 173–206.

Darling-Hammond, L., & Richardson, N. (2009). Teacher learning: What matters. *Educational Leadership*, *66*(5), 46–53.

Gallimore, R., Ermeling, B. A., Saunders, W. M., & Goldenberg, C. (2009). Moving the learning of teaching closer to practice: Teacher education implications of school-based inquiry teams. *Elementary School Journal*, *109*(5), 537–553.

Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teacher College Record*, *109*(4), 877–896.

Goddard, Y., Miller, R., Larsen, R., Goddard, G., Jacob, R., Madsen, J., & Schroeder, P. (2010). *Connecting principal leadership, teacher collaboration, and student achievement*. Paper presented at the American Educational Research Association Annual Meeting, Denver, CO.

Goldrick, L., Osta, D., Barlin, D., & Burn, J. (2012). *Review of state policies on teacher induction*. Retrieved from http://www.newteachercenter.org/sites/default/files/ntc/main/resources/brf-ntc-policy-state-teacher-induction.pdf.

Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal*, *48*(5), 1091–1123.

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher*, *42*(8), 433–444.

Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The test matters: The relationship between classroom observation scores and teacher value-added on multiple types of assessment. *Educational Researcher*, *43*(6), 293–303.

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making* (NCEE 2009-4067). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides.

Horn, I. S., & Little, J. W. (2009). Attending to problems of practice: Routines and resources for professional learning in teachers' workplace interactions. *American Educational Research Journal*, *47*(1), 181–217.

Jackson, C. K., & Breugmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, *1*(4), 85–108.

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, *36*(4), 476–500.

Lee, V. E., & Smith, J. B. (1996). Collective responsibility for learning and its effects on gains in achievement for early secondary school students. *American Journal of Education*, *104*(2), 103–147.

Levine, T. H., & Marcus, A. S. (2010). How the structure and focus of teachers' collaborative activities facilitate and constrain teacher learning. *Teaching and Teacher Education*, *26*(3), 389–398.

Little, J. W. (2002). Professional community and the problem of high school reform. *International Journal of Educational Research*, *37*(8), 693–714.

Little, J. W. (2003). Inside teacher community: Representations of classroom practice. *Teachers College Record*, *105*(6), 913–945.

Loeb, S., Kalogrides, D., & Beteille, T. (2012). Effective schools: Teacher hiring, assignment, development, and retention. *Education Finance and Policy*, *7*(3), 269–304.
Lortie, D. C. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago Press.
Louis, K. S., & Marks, H. M. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. *American Journal of Education*, *106*(4), 532–575.
Louis, K. S., Marks, H. M., & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal*, *33*(4), 757–798.
MetLife Foundation. (2009). *The MetLife survey of the American teacher: Collaborating for student success*. Retrieved from http://files.eric.ed.gov/fulltext/ED509650.pdf.
National Staff Development Council. (2001). *Standards for professional learning*. Retrieved from http://learningforward.org/standards/learning-communities#.UogfCGTk_R0
Papay, J., & Kraft, M. (2011). *Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career growth*. Unpublished working paper.
Saunders, W., Goldenberg, C., & Gallimore, R. (2009). Increasing achievement by focusing grade-level teams on improving classroom learning: A prospective, quasi-experimental study of Title 1 schools. *American Educational Research Journal*, *46*(4), 1006–1033.
Strahan, D. (2003). Promoting a professional collaborative culture in three elementary schools that have beaten the odds. *The Elementary School Journal*, *104*(2), 127–146.
Supovitz, J. A. (2002). Developing communities of instructional practice. *Teachers College Record*, *104*(8), 1591–1626.
Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, *24*(1), 80–91.

Manuscript received May 16, 2014
Final revision received March 28, 2015
Accepted April 13, 2015