Educational Evaluation and Policy Analysis
December 2018, Vol. 40, No. 4, pp. 586–612
DOI: 10.3102/0162373718785764
Article reuse guidelines: sagepub.com/journals-permissions

Constructing "Experts" Among Peers: Educational Infrastructure, Test Data, and Teachers' Interactions About Teaching

James P. Spillane
Northwestern University
Matthew Shirrell
George Washington University
Samrachana Adhikari
Harvard University

Teachers' on-the-job interactions with colleagues impact their effectiveness, yet little research has explored whether and how teacher performance predicts these interactions. Drawing on 5 years of social network data from one school district, we explore the relationship between teacher performance and teachers' instructional advice and information interactions. Results demonstrate that higher performing teachers are not more likely to be sought out for advice; instead, higher performing teachers are more likely to seek advice. Although school staff report they can identify the "best" teachers, they generally do not rely on student test scores, instead relying on more readily accessible indicators of performance. These findings have important implications for policy and practices that seek to promote desired interactions among teachers.

Keywords: *mixed methods, teacher assessment, teacher knowledge, testing*

Professionals can learn and improve their performance through their interactions with peers (Eraut & Hirsh, 2010; Leana & Pil, 2006; Pil & Leana, 2009; Supovitz, Sirinides, & May, 2010). Although not all such interactions enable teacher learning (Achinstein, 2002; Coburn & Russell, 2008; Grossman, Wineburg, & Woolworth, 2001; Hargreaves, 2000), teachers' peer interactions can, under the right circumstances, enable learning about teaching (Coburn & Russell, 2008; Horn, 2005; Little, 1990, 1993; Parise & Spillane, 2010; Smylie, 1995). Recent research demonstrates that lower performing teachers can benefit from interacting with higher performing peers in ways that improve their students' performance. One study, for example, shows that teachers' performance increases when they work alongside more effective colleagues, with 20% of what is commonly thought of as an individual teacher's "value added" attributable to that teacher's peers (Jackson & Bruegmann, 2009). Another study demonstrates that teacher performance improves when a more effective teacher joins a grade-level team (Sun, Loeb, & Grissom, 2017). Understanding what might account for which peers teachers interact with about instruction, then, is important.

Recent work identifies several predictors of these interactions, finding that teachers' individual characteristics such as race, gender, and instructional beliefs (Coburn, 2001; Spillane, Hopkins, & Sweet, 2015; Spillane, Kim, & Frank, 2012), aspects of the educational infrastructure such as formal position and grade-level assignment (Coburn & Russell, 2008; Spillane et al., 2015; Spillane et al., 2012), and even schools' physical infrastructures (Spillane, Shirrell, & Sweet, 2017) all predict the likelihood of teacher

interactions about instruction. Indeed, some school systems are intentionally redesigning their educational infrastructures to influence with whom and how teachers interact about instruction, focusing on aspects of the infrastructure such as teacher leadership and coaching positions, and creating structured time for teacher collaboration (Coburn, Mata, & Choi, 2013; Coburn & Russell, 2008; Penuel, Sun, Frank, & Gallagher, 2012; Spillane, Shirrell, & Hopkins, 2016; Sun, Penuel, Frank, Gallagher, & Youngs, 2013; Sun, Wilhelm, Larson, & Frank, 2014; Supovitz, 2006). Our study is based on data from one such district that designed and implemented an educational infrastructure to support teachers' interactions about mathematics instruction as they implemented a new inquiry-oriented math curriculum—a core and anchoring component of the infrastructure.

Another predictor of workplace interactions is the expertise of the individuals being sought out for advice. Research suggests that individuals seek advice from those whom they perceive as having expertise (Adler & Kwon, 2002; Coburn, Choi, & Mata, 2010; Coburn et al., 2013; Coburn & Russell, 2008; Penuel, Riel, Korbak, & Means, 2004; Rivera, Soderstrom, & Uzzi, 2010; Spillane, Hallett, & Diamond, 2003). Given that teachers' interactions with higher performing peers lead to improvements in teachers' own performance (Jackson & Bruegmann, 2009; Sun et al., 2017), understanding *how* teachers identify expert peers is crucial to creating school work environments that enable interactions about teaching.

There are several plausible ways that teachers might identify expert teachers among their peers. One possibility is that teachers use their peers' social status to identify expertise, assuming that peers who are more sought out for advice are experts (Bridwell-Mitchell & Cooc, 2016). Teachers may also use assignment to "formal" organizational roles such as leadership positions, participation in formal training such as professional development, or degree attainment as indicators of expertise (Spillane et al., 2003; Spillane et al., 2015). A third possibility is that teachers rely on student performance as an indicator of teacher expertise; in particular, as school systems focus increasingly on using student test scores to evaluate teacher performance (Donaldson & Papay, 2015; Jackson, Rockoff, & Staiger, 2014; Steinberg & Donaldson, 2016), this test data may have become a more salient indicator of expertise. Although research documents that perception of expertise influences with whom teachers interact about instruction, we found only one study showing that teachers are more likely to seek advice from colleagues based on their students' achievement gains (Wilhelm, Chen, Smith, & Frank, 2016).

In this article, we explore how teachers identify experts among their peers and whether and how teachers use the test performance of their colleagues' students in these determinations. Using a longitudinal mixed methods case study of one local school district that was working to shape teachers' interactions about instruction, we first analyze interviews to explore the ways that teachers think about the best teachers of mathematics among their colleagues, paying particular attention to the ways teachers describe the role of student test data in these constructions of teaching expertise. We then use social network analysis to examine whether measures of the test performance of teachers' students predict teachers' instructional advice and information seeking behavior. Our analysis explores how teachers use various types of information to construct their ideas of the most expert mathematics teachers among their peers; whether these constructions include the test performance of teachers' students; and whether or not student test performance predicts teacher interactions about instruction. We focus on mathematics for two reasons: first, district reform efforts were focused on mathematics in the district we studied; and second, other work suggests that teachers' interaction patterns about their teaching differ depending on the school subject (Spillane & Hopkins, 2013).

We argue that while school staff report knowing the best mathematics teachers among their peers, they generally do not rely on student test scores to identify those experts. Instead, in identifying experts they rely on their own direct or indirect knowledge of their colleagues' teaching approaches; their colleagues' excitement about, knowledge of, and understanding of mathematics, as demonstrated in meetings; and their colleagues' subject-specific formal positions and/or training. Furthermore, we find that various

components of the school district's educational infrastructure to support elementary math education created opportunities for teachers to build this direct and indirect knowledge about colleagues' math expertise. Consistent with our qualitative analysis, we find that, despite an extensive infrastructure focused on state and local tests, school staff were *not* more likely to seek out colleagues for instructional advice when those colleagues' students performed better on tests; instead, we find that higher performing teachers were more likely to *seek* advice and information than were lower performing teachers. We conclude by discussing the implications of our findings for policy, practice, and research.

Theoretical and Empirical Anchors

To ground our work, we first review the literature on the factors associated with teachers' instructional advice and information seeking behavior. We then review work on the role of instructional expertise in predicting teacher peer interactions, as well as work on teachers' and school leaders' understandings of student test scores as measures of teacher performance.

Predictors of Peer Interactions in the Workplace

Because social interactions are neither "a natural given" nor "a social given" (Bourdieu, 1986, p. 249), it is essential to identify those factors associated with the presence (or absence) of peer interactions, to better understand what might account for teachers' interactions with one another (Coburn, 2001; Small, 2009; Spillane et al., 2012). Prior work suggests that individuals tend to interact with those that are similar in terms of age, race, gender, education, and values, a finding commonly referred to as "homophily" (Feld, 1982; Ibarra, 1992; Leenders, 1996; Marsden, 1987; McPherson, Smith-Lovin, & Cook, 2001; Mollica, Gray, & Trevino, 2003; Monge & Contractor, 2003; Shrum, Cheek, & Hunter, 1988). Recent work on schools suggests that individual characteristics—in particular gender, race, and age—influence interactions among school staff (Frank & Zhao, 2005; Moolenaar, Sleegers, & Daly, 2012; Spillane et al., 2015; Spillane et al., 2012). In one recent study, for example, similarity of race and gender were associated with instructional advice and information interactions in schools (Spillane et al., 2012).

At the same time, organizational factors are also important to interpersonal interactions in the workplace. This is to be expected, as it has long been theorized that the formal organizational structure can enable and constrain interactions among its members (Blau, 1955; Blau & Scott, 1962). In education, several recent studies suggest that organizational factors do indeed matter to teachers' interactions, and have a larger influence than individual factors on the likelihood of such interactions (Spillane et al., 2015; Spillane et al., 2012). Specifically, teachers in the same grade level are more likely to interact with one another about teaching (Frank & Zhao, 2005; Moolenaar, Karsten, Sleegers, & Daly, 2014; Penuel, Frank, & Krause, 2010; Spillane et al., 2012; Sun et al., 2013; Wilhelm et al., 2016), and teachers that teach multiple grades are less likely to seek and provide advice and information (Spillane et al., 2016). Similarly, having a formal leadership position in the school positively predicts providing advice and information to colleagues (Moolenaar et al., 2012; Spillane et al., 2015; Spillane et al., 2012; Wilhelm et al., 2016).

Expertise, Student Test Scores, and Peer Interactions

Although teachers who interact with peers whose students perform higher improve their own students' performance (Jackson & Bruegmann, 2009; Sun et al., 2017), we know less about whether student test performance is a measure of expertise that teachers consider when determining whom among their colleagues to seek out for instructional advice. We found only one study that directly examined the role of student test data in teachers' instructional advice seeking behavior. This study, involving 109 middle school math teachers in 27 schools, examined whether five types or measures of expertise related to mathematics instruction (i.e., years of teaching experience, capacity to enact inquiry-oriented instruction, mathematical knowledge for teaching, instructional vision, and ability to improve student achievement) predicted whom teachers sought for mathematics instructional advice (Wilhelm et al., 2016). The researchers

found that teachers were more likely to seek advice from colleagues whose student achievement gains were better than their own, but the other four forms of expertise did not predict who teachers sought out for advice.

Still, research does suggest that perception of expertise is often a reason given for seeking someone out for advice (Adler & Kwon, 2002; Rivera et al., 2010). Several studies of teacher interactions show that teachers reach out to peers when they perceive them to have expertise (Coburn et al., 2010; Coburn et al., 2013; Coburn & Russell, 2008; Penuel et al., 2004; Spillane et al., 2003). One study involving 84 teachers in eight elementary schools in a large urban district, for example, found that teachers construct other school staff as instructional leaders that influence their teaching based on an assessment of their instructional expertise, among other factors (Spillane et al., 2003). Specifically, in discussing peers who influenced their teaching, teachers referred to their instructional knowledge, skills, and expertise as the basis for their influence (Spillane et al., 2003). Another study, involving four elementary schools in a mid-sized urban school district, found that one reason that teachers gave for reaching out to particular peers for instructional advice was the peers' knowledge and expertise in mathematics and mathematics instruction (Coburn et al., 2010; Coburn et al., 2013). Furthermore, the same studies showed that, over time, expertise became a more important basis for interacting with peers, due to school district efforts to reform mathematics instruction that included the creation of formal structures that enabled teachers to interact with one another (Coburn et al., 2010; Coburn et al., 2013).

Although the above-mentioned studies document the role of teaching expertise in shaping who teachers reach out to for instructional advice, this work (with the exception of one study) provides no evidence that student performance, as measured by test scores, is a means by which teachers assess their colleagues' expertise. Instead, these studies generally find that teachers refer to their peers' instructional performance more broadly. Work on teachers' and school leaders' perceptions of student test data—particularly its use as a measure of teacher performance—offers some explanations as to why this may be the case. Teachers and administrators are often skeptical of measuring teacher performance using student test scores, expressing significant concern, for example, about the inclusion of student growth measures in teacher evaluation systems (Jiang, Sporte, & Luppescu, 2015). Principals and central office administrators in urban districts also express concerns about the timing, transparency, complexity, and perceived validity of teacher value-added measures of teacher performance, and place much greater weight on their formal observations of teachers when conducting evaluations (Goldring et al., 2015). In evaluating teachers, principals instead value teacher characteristics such as effort, experience, and sociability (Harris, Ingle, & Rutledge, 2014). This evidence suggests that teachers and school leaders have mixed views about using student test scores to measure expertise, and suggests one reason that teachers may not rely on these measures when determining whom to talk to about their teaching.

Research Questions

To explore the ways that teachers construct their colleagues' expertise, the role of student test scores in these determinations, and whether or not student test performance predicts teacher interactions about instruction, we examine three research questions:

Research Question 1: What information do teachers and school leaders draw on when constructing their understandings of the most expert or "best" teachers of mathematics among their colleagues?

Research Question 2: How, if at all, do the test scores of their colleagues' students figure into these constructions of the "best" teachers of mathematics?

Research Question 3: Does teacher performance, as measured by their students' test scores, predict being sought out or seeking work-related advice about mathematics?

We answer the first two questions based on an analysis of qualitative interview data, and answer question three using student test score and teacher social network data.

Method

Our analysis is based on data from a longitudinal, mixed methods study that used a sequential explanatory mixed methods design (Creswell & Clark, 2011), whereby we implemented the quantitative portion of the study (i.e., survey), then selected participants for qualitative interviews based on their survey responses. Below, we first describe our case study site, then our research design, beginning with the qualitative component.

Study Site and Case Study

Auburn Park (pseudonym) is a mid-sized suburban district enrolling approximately 6,000 students in 14 elementary schools, of whom roughly 80% were White, 5% Latino/a, and 5% African American at the time of our study; 25% of students received free or reduced-price lunches. Auburn Park is a case of a school district working to reform mathematics instruction to support a more inquiry-oriented approach to elementary mathematics teaching. To do so, the district leadership committee for mathematics, composed of the district curriculum director and teachers, selected a new inquiry-based mathematics curriculum for elementary schools. Not leaving the implementation of the new curriculum to chance, the committee designed an elaborate educational infrastructure to support teachers in teaching the new curriculum.

Educational Infrastructure. This educational infrastructure was made up of several interrelated components, including revised district-level student assessments for mathematics, professional development in the form of a master's program for teacher leaders strategically selected from each elementary school, new mathematics coach positions in some elementary schools, school and system level organizational routines including weekly team and Professional Learning Community (PLC) meetings to support mathematics instruction, and an online system where teachers could access the performance of their colleagues' students on both state and district tests. District leaders designed and implemented these educational infrastructure components to work together to influence school staff interactions about mathematics instruction (Spillane, Hopkins, & Sweet, 2017).

Instructional expertise, including student tests and test scores, was an important component of the educational infrastructure. First, the district invested in developing the expertise of teacher leaders, carefully selecting expert mathematics teachers to participate in a master's program in mathematics and mathematics instruction, and hiring several of these teachers as either math coaches or math teacher leaders in their schools. Second, organizational routines such as grade-level PLCs were sites for developing expertise through interactions among staff with coaches and teacher leaders about implementing the math curriculum (Spillane et al., 2016).

Student mathematics assessment and related data featured in the educational infrastructure in at least two ways. First, teachers jointly scored district mathematics assessments and deliberated on the results as part of their weekly PLC and grade-level team meeting routines (Spillane et al., 2016). All interviewees, without prompting, referenced grade-level PLCs as a primary site for interacting with colleagues about instruction, and the vast majority of teachers we interviewed talked about scoring students' answers to district assessments together during these meetings. Second, the district created an online system where teachers could access their colleagues' students' performance in mathematics on both state and district tests, making teacher performance as measured by student test data readily available to school staff. Emma, the principal at Chamberlain captured how the educational infrastructure supported teachers' use of student test data:

> Our district has a data dashboard—and we are able to look at cumulative results and then we can look at subgroups and we can drill down to Special Education, male/female, free and reduced lunch, you know ethnicity . . . we look at program and cohort data and based on program and cohort data we'll set our goals for the following year . . . then quarterly the SIP [School Improvement Planning] team, our school improvement team, will go through and update all of those factors and share that with the staff. Now, with assessment sharing we will do that in PLCs. We double score in PLCs and then we also compare—you know we'll say in PLCs "why did this happen in this classroom but not in this classroom? Let's talk about that. What are you doing? What made your kids so

> successful here?" or "I see a pattern of success with data and probability but consistent lower scores in geometry and measurement. Why are girls doing better with number sense and algebra than boys?" . . . kind of hypothesis of why that's happening and what are we going do about it: based on this data what are we responsible for? . . . that transparency and the brutal facts and looking at data and having it out in the open.

Emma identifies two aspects of the district's educational infrastructure here that foregrounded student test score data: the district's "data dashboard" and the Professional Learning Communities; she also captures how these elements were integrated into other aspects of the educational infrastructure.

Auburn Park as a Case of. From a research perspective, Auburn Park provides a case of a school district working to reform mathematics instruction system-wide through the development of an educational infrastructure that was designed to influence who and what teachers talked about with respect to mathematics. At the time of our study, the development of an educational infrastructure to support math teaching and learning was the main policy initiative undertaken by the district; no other policies of similar magnitude took effect during the time of our study. Auburn Park is a relatively high-performing district, which did not face urgent pressures for improvement or to close achievement gaps, in a state that was among the last in the nation to introduce state standards and state assessments for both English/Language Arts and mathematics. Considering the prominence of student math assessments and assessment data in Auburn Park's educational infrastructure, we next explored whether and how they figured in teachers' constructions of expert math teachers among their peers.

Given our case study approach, we make no claims for statistical generalizability (Yin, 2005); instead, we think about external generalizability in two ways. First, in terms of "transferability," we argue that the applicability of our findings to other school districts should take into account the similarities and differences between these contexts and the Auburn Park setting we examine here (Lincoln & Guba, 1985). Second, we argue that our findings have "theoretical" or "analytical" generalizability, as our claims pertain to a *process* (as distinct from characteristics of a site) that is generalizable beyond our particular study site and that contributes to the refinement of theory, in this case theory about teacher peer interactions (Becker, 1990; Eisenhart, 2009; Small, 2009).

Qualitative Data and Analysis

Data. In spring 2015, we conducted semistructured interviews with 32 school and district staff in Auburn Park. We selected five schools using a purposeful sampling strategy (Lincoln & Guba, 1985; Patton, 1990) to maximize variation on dimensions believed important to school staff interactions about mathematics instruction, such as the presence or absence of a mathematics coach. Within each school, we also purposefully sampled staff to maximize variation in formal position, social network position (e.g., more or less central in their schools' mathematics network), and level (primary or upper grade teachers). We sought to interview a broad range of staff, to maximize opportunities for verification (or contradiction) of our theorizing about school staff interactions about mathematics instruction.

To ensure comparable data were collected across sites, we used a semistructured interview protocol. The portions of the protocol most relevant to this article asked interviewees about whom they went to for advice and information about mathematics, why they went to these individuals, *and* if and how they could identify the "best" math teachers in their schools. Anticipating, based on our prior work, that interviewees were likely to reference their colleagues' expertise in broad and general terms, and wanting to better understand how they understood expertise, we asked, "In general, do you have ideas about who the best math teachers are in this school?" If they responded that they did, we followed up by asking, "How do you know they're the best teachers?" If respondents did not mention student test scores, we probed specifically about the role (if any) of student test scores in their thinking about the best teachers. Interviews lasted approximately 45 minutes and were recorded, transcribed, and imported to NVivo 10 for analysis.

Analytical Approach. We analyzed our interview data in four phases. In Phase 1, we coded all

TABLE 1

Descriptive Statistics on Math Networks in Auburn Park Elementary Schools, 2010–2015

	2010	2011	2012	2013	2015
Network density					
M	0.07	0.06	0.07	0.06	0.07
SD	0.04	0.01	0.01	0.01	0.01
Minimum	0.04	0.04	0.04	0.05	0.05
Maximum	0.18	0.08	0.10	0.09	0.10
Number of ties per node					
M	1.54	1.74	1.82	1.74	1.88
SD	0.50	0.49	0.41	0.33	0.34
Minimum	0.55	1.08	1.17	1.18	1.30
Maximum	2.45	2.61	2.39	2.29	2.40
Number of nodes in network					
M	23.71	28.07	26.64	27.79	29.00
SD	7.70	4.71	5.62	4.82	5.35
Minimum	8	18	17	18	19
Maximum	32	35	34	35	36
n, networks	14	14	14	14	14
n, nodes	332	393	373	391	406

the reasons interviewees gave for seeking out a colleague for advice about mathematics instruction. In Phase 2, researchers each coded a subset of the interviews for any references to teacher expertise and/or performance; the researchers then met to discuss emergent themes and to develop an initial coding scheme. The coding scheme that emerged included three broad categories: (a) references to testing or assessment data as a basis for determining exemplary math teachers; (b) references to other aspects of teachers' practices as a basis for determining exemplary math teachers; and (c) other references to teacher performance that did not fit into either of these emergent codes. In Phase 3, the same two researchers coded the entire data set, focusing particularly on identifying themes in teachers' discussions of the role of performance in determining exemplary teachers. The fourth and final phase of analysis involved creating a series of analytic memos to explore these themes and document their prominence.

Quantitative Data and Analysis

Survey Data. Our quantitative analysis drew on 5 years of surveys of teachers and administrators in Auburn Park's 14 elementary schools; these surveys were administered in the spring of 2010, 2011, 2012, 2013, and 2015. The surveys asked teachers and administrators about their teaching positions, backgrounds, and work experiences, and to list the colleagues from whom they sought instructional advice and information about mathematics. Respondents were first asked to provide a list of colleagues they sought out for instructional advice or information: "During this school year, to whom have you turned for advice and/or information about curriculum, teaching, and student learning?" Respondents could list up to 12 names. After providing a list of names, school staff were asked whether or not they had received advice and/or information about math from each colleague they listed. These survey items were previously validated (Pitts & Spillane, 2009) and used in a number of other studies of teachers' advice and information networks (see, for example, Spillane et al., 2015; Spillane et al., 2012; Spillane, Shirrell et al., 2017). In 2010 through 2015, our survey response rates were 81%, 95%, 94%, 94%, and 96%, respectively, well above the 70% response rate required for social network analysis (Kossinets, 2006; Wasserman & Faust, 1994).

Table 1 provides descriptive statistics on math networks in Auburn Park elementary schools. The density of math networks—or proportion of possible ties that were actually present—averaged 0.07 over the 5 years of our data; as can be seen in the middle panel of Table 1, teachers sought out an average of 1.7 colleagues for advice about math. The bottom section of Table 1 shows the numbers of school staff in math networks each year. Schools' math networks had about 27 teachers, on average, across the 5 years we examine here.

We also used responses to several other survey items as covariates in our analyses, all of which have been found to significantly predict advice and information ties in schools (Moolenaar et al., 2014; Spillane et al., 2015; Spillane et al., 2012). These items included respondents' reports on having a leadership position such as a principal, coach, or grade-level leader; their gender; and their years of experience in education. (As our sample was nearly entirely White, we were unable to include a covariate that measured teacher race.) Respondents' reports of the grade level(s) they taught were used to create an indicator for whether a respondent taught more than

one grade, which was also included as a covariate. At the pair (dyad) level, we also included a covariate for whether two teachers taught the same grade level, which has been shown to strongly predict instructional ties between teachers (Spillane et al., 2012), as well as a covariate for whether two teachers taught in the same school. In some analyses, we included a covariate for whether the members of the dyad were both new to education (in years 1–3 years in the field), as well as the walking distance (in feet) between their workspaces, which prior work suggests is also a significant predictor of ties (Spillane, Shirrell et al., 2017).

Student Test Score Data. To create measures of teacher performance based on teachers' students' test scores, we used student test score data provided by the district. Auburn Park provided student test scores on district-mandated math exams for eight school years (2007–2008 through 2014–2015), with students' test scores linked to a single teacher in each year. District-mandated math exams were administered at several points during each school year, and the results were used for a variety of school improvement efforts; the district exams were "low stakes," however, as test performance was not formally tied to rewards or sanctions for teachers or schools.

The state administered a new, "high stakes" math assessment for the first time during the 2010–2011 school year. The stakes of this new state exam were higher than those of the district tests, as scores from the state test were used to create school rankings which received a great deal of publicity, particularly in the test's first years. The state math test was also "high stakes" because the federal government used scores from this test to determine whether schools had made "Adequate Yearly Progress" as defined by the state under the No Child Left Behind Act (NCLB) (although not in the first year they were administered). The state test, however, was administered only to students in Grades 3 to 6, limiting the number of teachers for whom we could compute performance measures based on student scores on these state exams.[1]

Auburn Park created a robust infrastructure to allow teachers and administrators to access their own and their colleagues' students' performance on both district and state tests. Student performance on both the district and state tests were made available to teachers through an online system, created by the district, which allowed teachers to examine the scores of their own students, as well as those of their colleagues.[2] In addition, this performance data—particularly results of the district exams—were discussed widely among teachers in grade-level meetings and in other school routines. The district-level assessments were scored by teachers themselves, but were frequently "double scored" during grade-level PLC meetings, when teachers graded the work of other teachers' students and had the opportunity to learn about those students' performance (Spillane et al., 2016). The district did not compute teacher value-added scores based on either district or state exams during this period, so such scores were not available to teachers during the time of our study.[3] Although principal evaluation ratings of teachers were also centrally collected (in a separate computer system from student test scores), these scores were made available only to administrators and district human resources officials, not to teachers,[4] so it was unlikely that such scores would inform teachers' perceptions of their colleagues' performance. We were not provided access to these data for our analysis.

Performance Measures. Using student test scores on both district and state mathematics exams, we created three test-based measures of teacher performance. These measures each captured a different aspect of teacher performance, and were included separately in our models to explore whether there were differences in the degree to which they explained advice seeking and giving in schools. In some cases, we included individual performance measures in our models; in other analyses, we included the *difference* between the performance of the members of each dyad, to determine whether that difference predicted the likelihood of a tie between them.

Proportion of proficient students. For its district-mandated exams, Auburn Park set thresholds that students had to meet to be deemed "proficient." The state where Auburn Park is located set similar proficiency thresholds for the state exams. To create our first test-based teacher performance measure, we calculated the percentage of each teacher's students that were deemed "proficient" on the relevant exam each year.

Class average test score. Our second test-based measure of teacher performance was the average test score of a given teacher's students in math. To compute this average, we standardized students' scores on each exam by grade and year across the entire district, then averaged each student's standardized scores on all the exams he or she took that year to create a yearly math score for each student. We computed this measure separately for district- and state-mandated exams in each year. We then averaged these scores by class in each year to create a class average, again separately for district- and state-mandated exams.

Teacher value-added. The final performance measure we created was an estimate of teacher value-added for each teacher. These scores were calculated separately for district and state exams for each teacher, using both teacher random effects and teacher fixed effects. We computed two sets of teacher value-added estimates to include in our analyses: One set of estimates used 2 years of data (from the focal year and the prior year) to estimate teacher value-added (e.g., the 2013 value-added calculations were based on student test scores from 2011–2012 and 2012–2013). A second set of value-added estimates were calculated using student test scores from the focal year and all available prior years (e.g., the 2013 value-added estimates were calculated using student test scores from 2007–2008 through 2012–2013). Appendix A (available in the online version of the journal) provides more detail on our methods of value-added estimation.

Year-by-year value-added estimates are preferable for our analyses, as the yearly estimates are each based on the same number of years of data, and therefore have more consistent standard errors across years. Value-added scores estimated using the second method, in contrast, have widely variable standard errors because scores in later years are estimated using more years of student test scores than in earlier years. Year-by-year value-added scores are also preferable for our analyses because they are more consistent with our other test-based performance measures, which also rely on the 1-year lag of student test scores. As our results did not differ significantly depending on number of years of data used to calculate value-added, we report results here from models that use the year-by-year estimates.

As described above, teacher value-added was not computed by the Auburn Park district, nor was it computed by the state, so school staff did not have access to their colleagues' value-added scores at the time of our study; teacher value-added scores, therefore, could not have directly influenced teachers' advice-and-information-seeking behavior. Teacher value-added, however, attempts to remove the impact of student sorting from estimates of teacher performance (Jackson et al., 2014); therefore, we include value-added to determine whether student sorting could have accounted for any relationships we observe. Just as value-added measures control for the composition of teachers' classes to isolate teachers' contribution to their students' performance, it is possible that teachers, perhaps intuitively, took into account the prior achievement or other characteristics of the students assigned to their colleagues when they judged their colleagues' effectiveness. Value-added measures of teacher performance, then, can be thought of as measures of teacher performance that take this possibility into account.

Given that value-added estimates are best able to distinguish between the highest and lowest performers (Jackson et al., 2014), in some analyses we include indicators for whether a staff member had very high or very low value-added (in the top quartile or quintile of the distribution), to examine whether such nonlinear measures of teacher value-added are associated with advice seeking and giving in schools. We divided teachers into these subgroups as opposed to including a squared form of the continuous performance measure because high value-added teachers are of particular interest from a policy perspective, due to their particularly significant impacts on student outcomes (Chetty, Friedman, & Rockoff, 2014); quartiles of the value-added distribution are also more easily interpretable than a squared term.

As we had test scores from district exams for each year of our study, we were able to examine the influence of performance on these district tests for all teachers in each year (2010, 2011, 2012, 2013, and 2015). Given that the state mathematics exam was introduced in 2010–2011, and our analyses focused on the impact of *prior year* performance on advice interactions, we were able to examine the associations between teacher performance calculated using results of state math exams and advice interactions in 3 years

TABLE 2

Descriptive Statistics on Sample, Auburn Park, 2014–2015

	(A) Teachers with state test data	(B) Teachers with district test data
Teacher characteristics		
Taught multiple grades	0.15	0.15
Held leadership role	0.08	0.07
Female	0.82	0.87
Years of experience	10.42	10.00
	(8.13)	(8.30)
Performance measures		
Lag percent proficient, math (state)	0.82	
	(0.01)	
Lag math score (state)	0.59	
	(0.06)	
Lag percent proficient, math (district)		0.90
		(0.08)
Lag math score (district)		0.78
		(0.04)
Math value-added (random effect)	−0.045	−0.016
	(0.95)	(1.00)
Math value-added (fixed effect)	0.03	0.05
	(1.00)	(0.94)
Teachers, *n*	88	129

TABLE 3

*Correlations Among Math Performance Measures, Auburn Park, 2014–2015 (*n *= 88)*

		(1)	(2)	(3)	(4)	(5)	(6)
(1)	Lag percent proficient, math (district)	1.00					
(2)	Lag percent proficient, math (state)	.37	1.00				
(3)	Lag math score (district)	.80	.31	1.00			
(4)	Lag math score (state)	.36	.75	.39	1.00		
(5)	Math value-added (random effect)	.32	−.07	.56	.09	1.00	
(6)	Math value-added (fixed effect)	.10	−.19	.28	−.09	.25	1.00

Note. Sample limited to teachers that have all measures. Darker colors signify higher correlations.

(2012, 2013, and 2015), and only for teachers of Grades 3 to 6. Teachers included in the analyses of state math test performance were similar to teachers included in our analyses of district math test performance (see Table 2).

Correlations between performance measures. Correlations between the performance measures included in our analyses were generally moderate to small (Table 3). As might be expected, the highest correlations were between average math scores and percent proficient on both state and district tests.

Analytic Methods. As social ties are inherently interdependent and thus violate the assumption of

independence made by most regression methods, social network analysis requires particular methods (Borgatti, Everett, & Freeman, 2014; Wasserman & Faust, 1994). We use "latent space network models," which "control" for dependence among observations by including latent space positions between each pair of nodes in the network (Hoff, Raftery, & Handcock, 2002). These latent space positions—which can be thought of as akin to the residuals in a regression model—account for dependence structures in the network, including reciprocity, transitivity, and clustering, among others (Hoff et al., 2002; Sweet, Thomas, & Junker, 2013). As in other social network methods, latent space models categorize those in the network (commonly referred to as "nodes") as either "senders" or "receivers," depending on the direction of the ties between them. The model then estimates the likelihood of a tie between actors in a network based on covariates at the pair (dyad), sender, receiver, and network levels (Sweet et al., 2013).[5]

To take advantage of the longitudinal nature of our network data, we use a variation of latent space network models which adapt the latent space concept for use with longitudinal data by additionally controlling for the temporal correlations between networks over time. The models we use (Temporal Latent Space models) account for two distinct aspects of network structure, one cross-sectional and the other longitudinal. Cross-sectionally, our models control for the ways that network structures affect the tendencies of individuals to form ties with one another—for example, for the tendency of school staff to seek advice from "friends of friends." Longitudinally, these models also control for the tendencies of entire network structures to persist to varying degrees from year to year, and thus to be correlated with one another over time. In contrast to other network models commonly used with cross-sectional and longitudinal network data (such as p-star models and simulation investigation for empirical network analysis [SIENA] models), the models we employ control for network structures both cross-sectionally (within years) and across years (longitudinally), and are thus particularly well-suited to isolating the associations between covariates of interest and ties.[6] By controlling for network structure via the latent space positions, however, the models are unable to estimate the impacts of network structure itself on ties (e.g., the degree to which transitivity and reciprocity predict tie formation). However, since the focus of our analysis was on examining the association between covariates of interest (test-based performance measures) and ties, latent space models were particularly well suited to our analysis.

In our Temporal Latent Space models, the dependent variable Y_{ijt} indicates the presence of a tie directed from staff member i to staff member j at time t. The models were estimated as follows:

$$logit\left(Pr\left(Y_{ijt}=1\right)\right)=\beta_0+\sum_{k=1}^{K}\beta_{kt}X_{ijkt}-\ Z_{it}-Z_{jt}$$
$$Z_{it}=Z_{i(t-1)}+\varepsilon_{it}$$
$$\varepsilon_{it}\sim Normal\left(0,\Sigma\right)$$
$$Z_{i1}\sim Normal\left(0,\Sigma\right)$$

Here, the Z_{it} and Z_{jt} are the latent space positions for staff members i and j in year t, and are very close to the latent space positions of the corresponding staff members in the previous year, which accounts for the temporal dependence over time. The X_{ijkt} are a set of node-, edge-, and network-level covariates, which include a measure of teacher performance (described above) from the prior year. We also include several other covariates as controls: For tie senders and receivers, we control for whether a school staff member had a leadership role (such as grade-level leader, mentor, or coach), taught multiple grades, and their number of years of experience in education (standardized by year); and at the dyad (pair) level, whether two staff members taught the same grade level, were the same gender, or taught in the same school. As discussed earlier, in some analyses, we additionally include measures of the similarity of experience between the dyad, as well as the walking distance between their workspaces in the school building.

Although teacher performance could impact which teachers are sought out for or seek advice, it is also highly plausible that ties could impact performance, as prior work on peer learning among teachers suggests (Jackson & Bruegmann, 2009; Sun et al., 2017). The general problem of making causal inferences about endogenous social effects—what Manski (1993) calls "the reflection problem"—is particularly acute in

social network settings (Shalizi & Thomas, 2011), and a variety of methods have been proposed to surmount this challenge, some focused on network settings, others not (see Cohen-Cole & Fletcher, 2008; Steglich, Snijders, & Pearson, 2010; Steiner, Cook, Shadish, & Clark, 2010). In our analyses, we address this challenge by using teacher performance from the prior year (or prior years, in the case of some of our value-added estimates) to predict whether teachers are sought out for or seek advice in the year that follows. Given that performance is measured before ties, it is reasonable to believe that we capture the impact of performance on ties, and not vice versa. Still, it is possible that ties from a prior period, which are potentially correlated with the ties we measure, could have influenced performance. Given the cross-sectional and observational nature of our data, as well as the potential for confounding variables that are not controlled for in our analyses, our estimates should not necessarily be interpreted causally.

Results

Our findings are organized as follows: We begin by addressing our first two research questions, exploring the role of expertise in reaching out to peers and theorizing how school staff constructed expert math teachers. We then take up our third research question, examining whether teacher performance as measured by student test scores predicts teachers being sought out for, or seeking, mathematics instructional advice, while controlling for other factors known to influence who teachers seek out for advice.

Using Information to Construct Understandings of the "Best" Teachers

Consistent with prior work, interviewees typically offered several reasons for reaching out to particular colleagues for advice about mathematics teaching. Along with teaching the same grade, these included a person's leadership position (e.g., math coach), formal training in mathematics, participation in an organizational routine (e.g., PLC), physical proximity, and friendship. Furthermore, and also consistent with prior work, 24 of the 32 interviewees offered a colleague's expertise as one of their reasons for reaching out to them for instructional advice about mathematics. For most, their expertise-related responses tended to be brief and somewhat vague; representative examples included "knowledgeable" (Rachel); "she's a great source of ideas" (Mary); "she's interested in new ideas" (Loretta); "knowledge and passion for math instruction" (Emma); "she has some more of that background knowledge" (Becky); "she's a brilliant mind to pick" (William); and "she just has a greater depth of math knowledge than I do" (Lois). In each of these cases, respondents did not reference specific evidence of their colleagues' expertise. To get a better sense of how interviewees constructed their colleagues as expert math teachers, then, we turned to how they identified the best teachers.

All school staff, when asked, reported being able to identify the "best" math teachers in their schools. When probed about *how* they identified exemplary mathematics teachers, most (29 of 32) referred to something *other than* student test scores. Specifically, school staff referenced one of three things: a teacher's mathematics instruction; a teacher's knowledge of, and disposition toward, mathematics; or a teacher's leadership position or formal training related to mathematics. We consider each of these below, and then explore how student test scores figured in teachers' thinking about exemplary mathematics teachers.

Mathematics Instructional Practice. Twenty-four of the 32 interviewees referred to teachers' mathematics instructional *practice* as a means of identifying the "best" math teachers in their schools. In particular, interviewees referenced these exemplary teachers' styles of questioning students, the organization and flow of their mathematics lessons, and their ability to generate student engagement, enthusiasm, and excitement for mathematics. Mary Beth, the principal of Chavez school, explained,

> My best math teachers ask the best questions and they're able to follow up with those "why" [questions] and they're able to, when kids respond maybe push their thinking even further. My best math teachers are able to identify what's good math thinking, and maybe what isn't, and with that information they're able to give kids feedback . . . it seems natural to them, it's a language that they use . . . it's fun to be in any of those rooms . . . math is fun. And I walk out of

> their classes thinking "wow, I feel like I've learned something." . . . kids are involved, kids are engaged, kids are participating. There is thinking going on, there's a lot of discussion.

Mary Beth's sense of exemplary teachers of mathematics was grounded in how they taught mathematics, including their ability to ask good questions, probe student responses to these questions in ways that push students' mathematical thinking, and engage students with mathematics as a language and way of thinking.

Others also pointed to being able to identify good mathematics teachers based on what they observed in classrooms and/or what they heard teachers say about their teaching and their students' learning. Emma, the principal at Chamberlain, noted, "I can't really say that the [student test score] data really does make me think that that's the best math teacher . . . It really is mostly me observing how they instruct more than the data." Emily, a mathematics coach at Bryant, noted,

> there's a flow to their lesson, . . . vertically and horizontally they know where their kids need to be . . . they can diagnose situations like if one kid is using this strategy they can see what, where they need to go with that strategy, they can see maybe where they need some help . . . Their questioning is just off the charts.

Emily's sense of who are the good mathematics teachers at Bryant is based on the instructional practices she observes in these teachers' classrooms, including their ability to diagnose students' mathematical thinking and use that diagnosis to determine their next pedagogical move. Similarly, Katie, the media specialist at Kingsley, explained how she recognizes exemplary mathematics teachers:

> a lot of it for me has to do with engagement; classrooms where you go into and the kids are engaged in what's being taught. . . . If they can really get their kids to sink their teeth into information, in an activity in what they're doing, that to me is a good teacher.

For Katie, then, student engagement with the mathematical content was an indicator of expertise in mathematics teaching.

Some school staff relied on their colleagues' accounts of how they taught mathematics rather than direct observation because, as several teachers remarked, they had relatively few opportunities to observe their colleagues' mathematics teaching. Sophia, a second-grade teacher at Bryant, reported identifying good mathematics teachers not from observing their teaching directly but from "listening to the way they talk about their students and the way they work with them [students]." Jessica, a Chamberlain fifth-grade teacher, argued that she knows the best mathematics teachers because:

> how they question the kids more to be able to get more reasoning out of them . . . try to get the kids to really write it out and think it out before just writing the numbers down . . . explaining how they did it. . . . Last year during that book review for the book study, for the math, I think that, it really came out then who really did a good job of teaching math.

Furthermore, some teachers noted how they know the best math teachers based on the mathematical abilities of the students they received from these teachers. Joanne, a fifth-grade teacher at Bryant, reported that "fourth-grade teachers do an excellent job of preparing kids; they come to fifth grade really ready for math." In these instances, school staff relied on either direct or indirect evidence of how a colleague taught mathematics to identify the "best" teachers of mathematics.

Knowledge of, and Enthusiasm for, Mathematics. Eighteen of the 32 interviewees referred to teachers' knowledge of mathematical content and pedagogy, their enthusiasm for mathematics, as well as their ability to explain mathematics to their peers as bases for identifying expert math teachers. Clarissa, a fifth-grade teacher at Kingsley, explained that she could identify exemplary mathematics teachers because they are

> the ones that you hear at staff meetings or at staff development days bring up something that's very valid that maybe "I never thought about before." Or the ones who understand what the data from the assessment means more and they can explain it . . . and be able to put it in common language and I'm like "oh, ok. Got that."

For Clarissa, it was not only colleagues' knowledge, but also their ability to explain something in everyday language, that made them experts.

Similarly, Becky, a Chamberlain fourth-grade teacher, explained that knowledge of good mathematics teachers

> comes through in staff meetings or when we meet together on some of that vertical teaming . . . not just the way they present it but some of the different topics they talk about and the strategies that they have. And the way that they bring ideas to the table . . . the background knowledge that they can bring and how they approach it . . . "well this student needs to have da-da-da-da to be proficient. They're lacking in this area. Make sure that you . . ."

In Becky's view, peers' knowledge, and in particular their pedagogical knowledge, was especially salient in identifying good mathematics teachers. Layla, a third-grade teacher at Ashton, explained that she knows the good mathematics teachers because, "at like staff meetings . . . they're more um, passionate about it [mathematics]. You can see when they talk about math that they kind of know what they're talking about; they're more confident with it." Courtney, the literacy facilitator at Chavez, explained that she knows the good mathematics teachers because "they feel comfortable about it when we talk about it at PLC's. They're not afraid of math." In sum, school leaders and teachers used their first-hand experiences of teachers' knowledge of, comfort with, and enthusiasm for mathematics in school meetings to gauge good mathematics teachers.

Formal Position and Formal Training. School staff also referenced their peers' formal positions (e.g., math coach) or their formal training (e.g., participation in professional development about mathematics) as evidence that they were good math teachers. In particular, several staff referenced "Fundamental Math," the master's program in mathematics and mathematics education offered by a local university in partnership with the school district that was designed to prepare mathematics teacher leaders. Laura, a third-grade teacher at Kingsley, explained that the good mathematics teachers in Auburn Park are

> Fundamental Math people. I really see them as the experts and that's how they've kind of been presented to us . . . by having those friends go through that Fundamental Math program I know . . . how intense it was and I know they all said it was a lot but it really changed their understanding and thinking about math . . . we don't get a ton of time to go in and observe another classroom to actually see you know math in action to say that that's why I think that. But I think overall knowing that they've gone through the program and they are advocates of math and exude a confidence about math, I would say that that's why I think they're the experts.

Here, Laura acknowledges that as a regular classroom teacher that she has very limited time to observe her colleagues' teaching to make a determination about exemplary mathematics teaching, so she relies on their participation in a professional development program co-sponsored by the district and a local university. Rachel, a Kindergarten teacher at Chamberlain, similarly explained,

> the ones that have been through the Fundamental Math program. It's so extensive . . . I think they're very, very knowledgeable. I definitely would go to any of them (chuckles) for help. Or with like our math coach . . . they're the strong suits for sure.

For Rachel, formal training was also important, as was holding a formal leadership position—school mathematics coach. School staff therefore reported identifying exemplary mathematics teachers based on proxies for expertise such as formal position and/or formal training.

Student Test Scores and Understandings of the "Best" Teachers

Fourteen interviewees referenced student test data as a basis for identifying exemplary math teachers, but in all but four cases, these references emerged only when prompted directly by the interviewer. No one reported relying solely on student test data to identify a colleague as an expert math teacher. Eloise, the principal at Kingsley, noted,

> we have our district assessments . . . we just got done doing the state testing . . . if I see that there's pretty good performance on that I'd like to say that we're pretty solid teachers in the area of math. But mostly district assessments, informal assessments that teachers might give; those types of things.

Emily, the math coach at Bryant, explained that she knows good mathematics teachers:

> because data shows it . . . district [tests], state test scores . . . I can drill down to any teacher. If we see

> something, if we see some data that's out of the norm for that grade level then we will go to the student level and the teacher level and say "you know these things are going on in this classroom."

Jillian, the principal at Bryant, also noted how she uses student test data selectively:

> maybe I do check it subconsciously against with like "ok, so is this saying the same thing?" I would say our data is pretty good here so . . . it's probably I use data more when I'm looking at teachers that are, and mediocre is a bad word, but that are average or that I'm concerned about. I'm using their data and looking at that much more closely. The rock star math instructors, like their data, I mean their data is going to show me that anyway and . . . I don't need the data to back up what I know I see and I feel and I hear. But those that are like on that borderline where I'm like "ooh, how worried should I be about what's going on here?" Then I probably rely on it a little bit more.

Jillian's account suggests that while she uses student test scores in determining teacher performance, it is more often to identify "mediocre" or "average" teachers rather than to identify the exemplary, "rock star" math teachers. Lois, the math coach at Stevenson, reported identifying the best teachers more by how they *used* their student test data, rather than on their scores. She noted about the "best" math teachers:

> when they get their results back they then ask "what could I have done differently?" or "what do I need to do differently for these kids?" So, they're reflective . . . they're always seeking out ways to help those students . . . that aren't proficient. . . . I think that test scores are a part of it but how the teacher reacts to those test scores and uses those test scores is a big piece to that puzzle.

Lois contrasted these teachers who reflected on their students' test scores in an effort to improve their practice with less-than-exemplary math teachers, whose response to these scores was that they could not have done anything differently.

School leaders' and teachers' weak attention to student test scores as an indicator of teacher expertise had several origins. One reason, as suggested by Jillian's response above, is that school staff may have seen student test scores as more suited to identifying weak or average teachers than exemplary teachers. Another reason, explicitly referenced by 12 interviewees, was a more general skepticism about using student test scores as a basis for determining good teachers. Some expressed skepticism about whether student test scores, especially from state assessments, were capturing the kinds of thinking and problem solving that teachers were trying to emphasize in their mathematics teaching. Others argued that teachers should not be judged on test scores because their students changed every year, and so a great deal of students' test performance was beyond the control of individual teachers. Loretta, a second-grade teacher at Chamberlain, explained how she did not know her colleagues' state test scores, but even if she did, she was not convinced that she would use these data to identify the best teachers:

> I don't think I would consider their state test scores. I've never really thought about that but, nope (chuckles). I mean I'm sure if they were like horrible I would kind of, but our school overall the state test scores are pretty high.

Loretta acknowledges that if test scores were very low, she might consider such scores in making a determination about the quality of a teacher's mathematics teaching but goes on to note that this is not an issue at her school. Instead, Loretta argues that it was more important to look at teachers' "understanding of how students build their understanding in that subject area and the value of conversations and how much time they allot to the conversation piece and the development of student understanding." Similarly, Sophia, a second-grade teacher at Bryant, explained that identifying good mathematics teachers involves:

> more than just data; I think there's a growth to be looked at as well . . . I think there are too many variables to do that. So, if you looked at strictly data and said "well this student can get the right answer but can they explain and understand that material?" I think so many more pieces go into that because there's that writing piece; can they write what they're thinking? . . . So, I don't think it's set in stone that the higher your test scores are, the better the teacher is. . . . I also think that there are other factors we, we have mobility right here that can change things too.

These interviewees were concerned that student math tests did not measure the sort of mathematical learning they valued (e.g., students' ability to explain their answers), and/or that scores on these tests could only be correctly interpreted

TABLE 4

Associations Between Prior Math Performance on District Tests and the Likelihood of Math Ties, Auburn Park, 2010–2015

	Covariate	2010	2011	2012	2013	2015
Provider	Lag percent proficient, math	−0.234	−0.086	**0.356***	0.011	−0.077
	Lag math score	−0.182	−0.044	**0.335***	0.027	−0.155
	Math value-added (random effect)	0.012	−0.096	−0.095	−0.125	**0.337***
	Math value-added (fixed effect)	0.072	−0.240	0.033	−0.119	0.480
Seeker	Lag percent proficient, math	**0.373***	**0.641***	**0.702***	**0.745***	**0.526***
	Lag math score	**0.432***	**0.663***	**0.705***	**0.811***	**0.542***
	Math value-added (random effect)	−0.306	−0.057	**–0.623***	−0.147	0.105
	Math value-added (fixed effect)	−0.134	**0.498***	**0.311***	−0.230	**–0.559***

Note. Significant values are marked in bold and with asterisks; significance determined by 95% credible interval not containing zero.

when taking into account various other factors that influenced student test results from 1 year to the next.

Overall, our analysis identified a number of ways in which school staff determined the best math teachers among their colleagues and how several of these ways of identifying experts were enabled by the district's educational infrastructure. For example, formal positions and training were important aspects of the educational infrastructure that staff took into account when determining the best math teachers. Organizational routines such as PLCs, grade-level team meetings, and district-level math meetings were also important to teachers' constructions of experts, as these routines gave teachers access to—and allowed them to identify—particular colleagues as experts based on their knowledge, accounts of their teaching practice, and passion for mathematics.

At the same time, student test data, which was another component of the district's educational infrastructure, did not feature prominently in interviewees' constructions of the best teachers of mathematics. Yet despite the limited attention to student test scores in school staff constructions of exemplary math teachers, it remained possible that these scores influenced decisions about whom staff sought for mathematics instructional advice; in other words, it was possible that school staff used test data, as Jillian, the principal at Bryant, put it, "subconsciously." School staff may not have been aware of relying on student test scores as a metric for their colleagues' expertise, for example, or may have felt it was not socially desirable to say so. Hence, we examined whether student test scores influenced who teachers actually turned to for mathematics instructional advice and information.

Student Test Scores as Predictors of Interactions About Math

District Test Scores. In Auburn Park, the educational infrastructure was designed to provide teachers with direct access to their colleagues' students' test scores; in addition, teachers met frequently to discuss test score data. Given their level of access to information on the test performance of their colleagues' students, as well as the general expectation that educational decision-making would be based at least in part on this information, it was reasonable to expect that teachers would use this information when determining who to seek for advice, and perhaps more frequently seek advice from colleagues whose students performed better on those tests. Our analyses, however, showed that teacher performance on district tests—no matter how such performance was measured—was generally not predictive of being sought out for advice about mathematics (i.e., being an advice "provider") the following year (Table 4, top panel); in other words, higher performing teachers were no more likely to be sought out for advice about mathematics than their lower performing peers. This was true whether performance was measured using the percentage of a teacher's students that

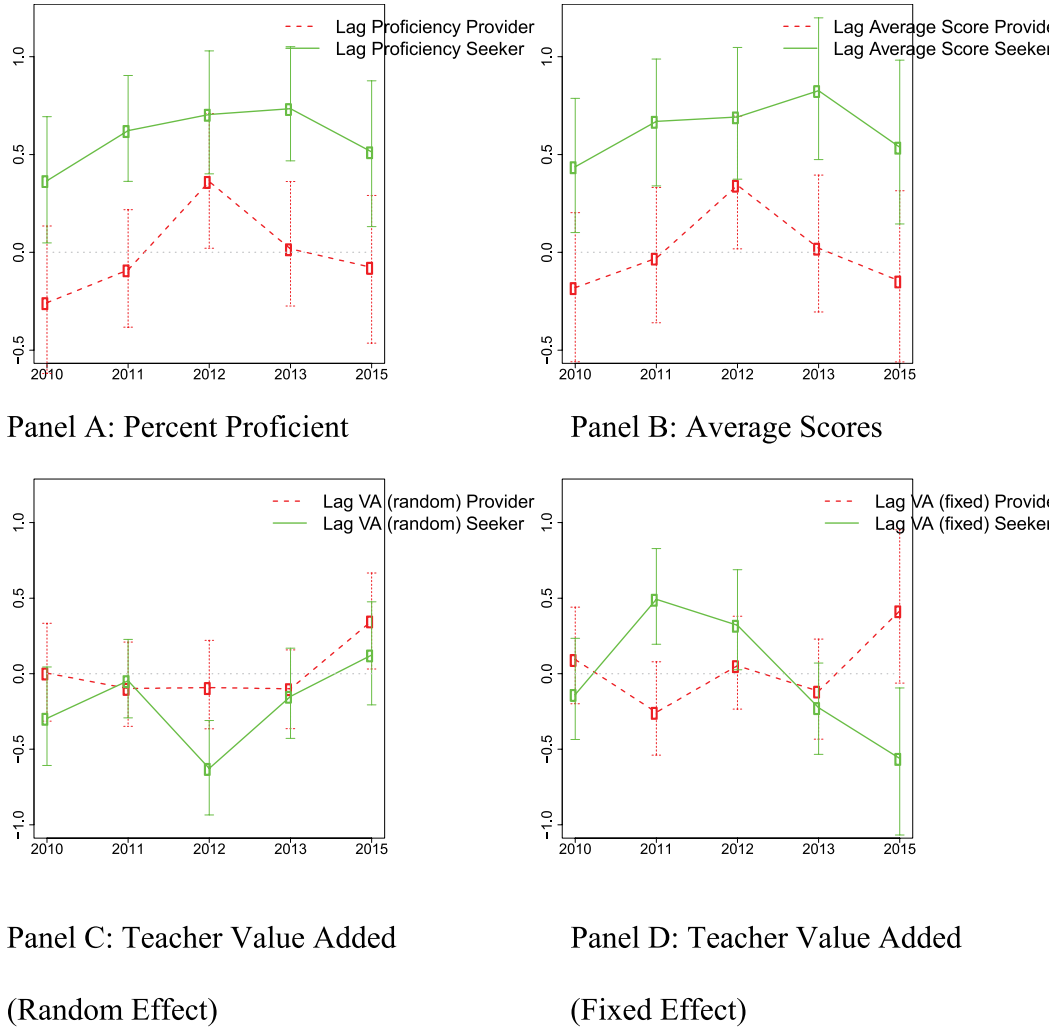

FIGURE 1. *Associations between various measures of teacher performance (calculated using district tests) and advice providing and receiving in math networks, Auburn Park, 2010–2015.*

were deemed proficient on the district tests the prior year, those students' average test scores, or teacher value-added to student achievement (Table 4, top panel).

The results of these analyses are depicted visually in Figure 1, where the points joined by the dashed lines represent each year's estimates of provider effects. The 95% credible intervals surrounding the estimates, which are represented by the dashed bars extending above and below the points, generally overlap with the horizontal line at zero, indicating that teachers whose students performed better on district math tests the previous year were *not* any more or less likely to be sought out for advice about math by their colleagues. As noted above, these analyses control not only for structural aspects of networks and the correlations between these structures over time but also for whether staff held a leadership role, taught multiple grades, and their gender and years of experience in education; at the dyad level, the analyses also control for whether the dyad taught the same grade level and worked in the same school. Although the estimates for these covariates are not shown here, teaching in the same school was the strongest positive predictor

TABLE 5

Associations Between Prior Math Performance on State Tests and the Likelihood of Math Ties, Auburn Park, 2010–2015

	Covariate	2012	2013	2015
Provider	Lag percent proficient, math	0.250	0.181	−0.291
	Lag math score	0.422	0.169	−0.401
	Math value-added (random effect)	−0.528	−0.449	−0.328
	Math value-added (fixed effect)	−0.005	−0.404	−0.518
Seeker	Lag percent proficient, math	**0.609***	**0.474***	**0.368***
	Lag math score	**0.992***	**0.621***	0.438
	Math value-added (random effect)	−0.273	0.399	0.119
	Math value-added (fixed effect)	−0.026	−0.057	0.430

Note. Significant values are marked in bold and with asterisks; significance determined by 95% credible interval not containing zero.

of ties in the models and, similar to the findings of prior work (e.g., Spillane et al., 2012), teaching the same grade level was also a strong positive predictor of ties.

Although the performance of a teacher's students on district math tests did not predict that teacher being sought out for advice, the performance of a teacher's students on those tests—when calculated as the percentage of proficient students or by students' average test scores—positively predicted *seeking* advice about math (Table 4, bottom panel). In other words, the better a teacher's students performed on these two measures (both based on the previous year's district math tests), the more likely that teacher was to *seek* advice from his or her colleagues about math the following year. Student performance on district tests, as measured using teacher value-added to student achievement, however, was not predictive of seeking advice (Table 4, bottom panel). This suggests that teachers with higher value-added were not any more or less likely to seek advice than were teachers with lower value-added.[7] These associations are depicted by the solid lines in Figure 1. In Panels A and B of the figure, the credible intervals (solid bars) surrounding the dots connected by the solid lines never cross the horizontal line at zero, indicating that across the 5 years we examine, teachers whose students performed better on district math tests the prior year were significantly more likely to *seek* advice from their colleagues. This association appears to climb slightly between the years 2010 and 2013 before declining somewhat in 2015; however, as the 95% credible intervals for each estimate overlap, it is impossible to statistically distinguish these estimates from one another. In Panels C and D of Figure 1, the solid estimates show no consistent pattern of difference from zero, indicating that teacher value-added (computed using scores from district tests) did not have consistent associations with advice seeking in any of the years we examine. Adding an indicator for both members of the dyad being new to education, as well as a measure of the walking distance between workspaces, did not significantly alter these results.

State Test Scores. Similar to our results for district tests, teacher performance as determined by students' scores on state math tests the prior year—no matter how performance was measured—was not associated with the likelihood of providing advice (Table 5, top panel). Teacher performance as measured by the percentage of proficient students and those students' average test scores on the state tests, however, was again associated with the likelihood of *seeking* advice (Table 5, bottom panel), similar to our findings for district tests, indicating that teachers whose students performed better on state tests were more likely to *seek* advice from their colleagues the following year. Teacher value-added (calculated using state tests) did not predict either being sought out for or seeking advice (Table 5). These results are depicted in Figure 2, where in the

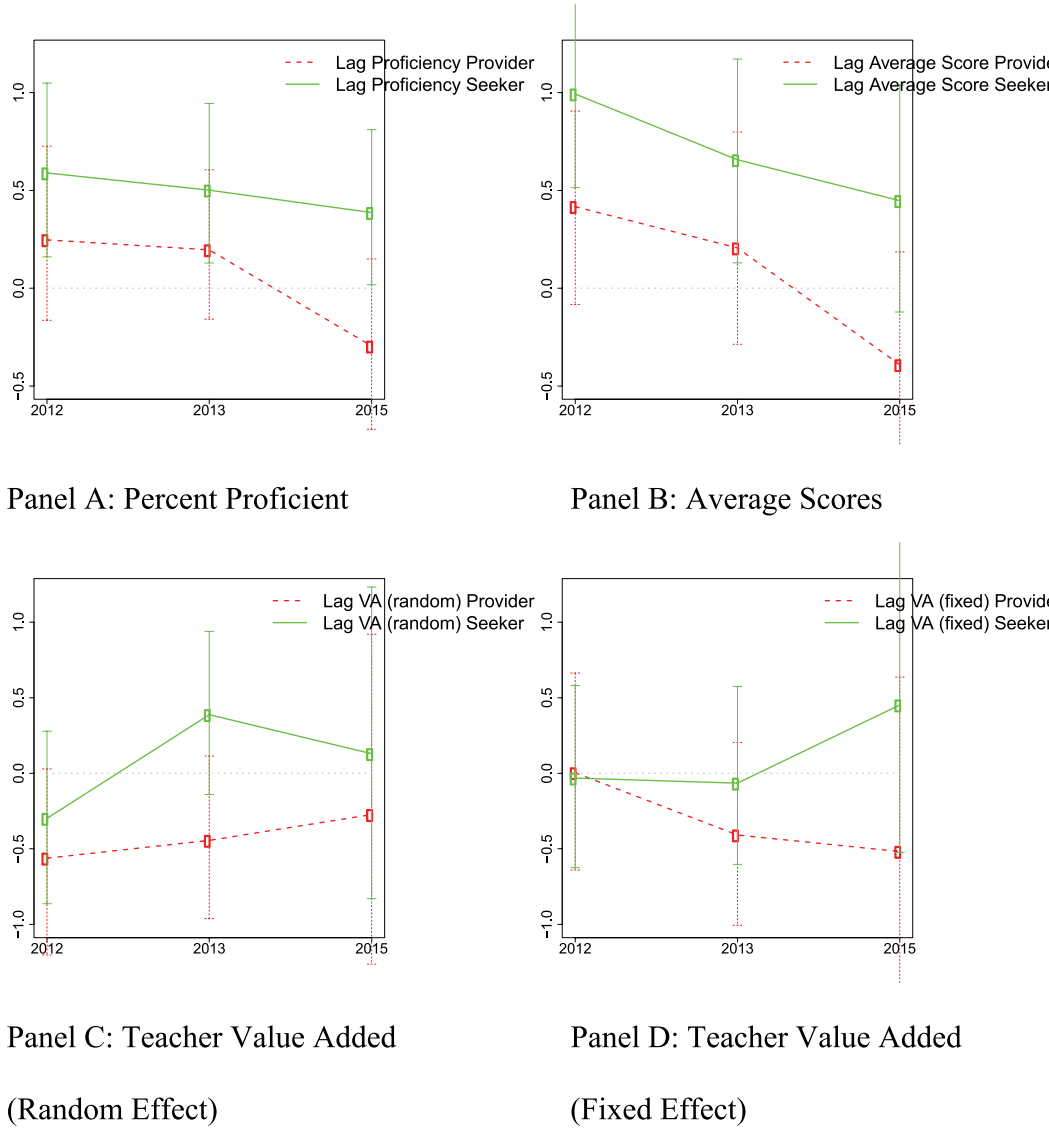

FIGURE 2. *Associations between various measures of teacher performance (calculated using state tests) and advice providing and receiving in math networks, Auburn Park, 2012–2015.*

top two panels the confidence intervals around the dashed estimates—which represent provider effects—generally overlap the horizontal line at zero, while the intervals around the solid estimates—which represent seeker effects—generally do not, again illustrating that teacher performance as measured by student scores on state tests was associated with advice seeking, but not with providing advice. The bottom two panels of Figure 2 illustrate that teacher value-added, here calculated using state tests, did not predict providing or seeking advice. Again, both members of the dyad being new to education or the walking distance between workspaces did not significantly alter these results.

Characteristics of Low and High Performers and Those They Sought for Advice. The preceding analyses show that the test performance of teachers' students did not predict teachers being sought

TABLE 6

Characteristics of Low and High Performers, Auburn Park, 2010–2015

	Low performers (bottom quartile)	High performers (top quartile)
Female	0.91	0.89
	(0.28)	(0.32)
Advanced degree	0.59	0.57
	(0.49)	(0.50)
Years experience in education	10.55	14.08**
	(9.08)	(9.05)
Years experience in school	6.93	9.37*
	(7.49)	(7.78)
Leadership role	0.17	0.22
	(0.37)	(0.42)
Multiple grades teacher	0.02	0.03
	(0.13)	(0.18)
Math coach	0.00	0.01
	(0.06)	(0.11)
Less than 4 hours math PD	0.46	0.45
	(0.50)	(0.50)
17† hours math PD	0.10	0.10
	(0.30)	(0.30)
n	253	261

Note. Analyses pool data from 2010, 2011, 2012, 2013, and 2015. High and low performers are determined separately for each year using the distribution of class average student math test scores on district tests. Tests of equality between columns account for repeated observations of some individuals. PD = professional development.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

out for advice about mathematics; such performance, however, was associated with *seeking* advice from others. Table 6 compares the characteristics of the lowest- and highest-performing teachers, using student test scores from district tests to compare teachers in the bottom and top quartiles of performance across the years 2010 through 2015. Relative to low performers, high performers had more years of experience in education and in their schools; however, high and low performers did not differ significantly in their gender or the likelihood they held an advanced degree, were a math coach, school leader, taught multiple grades, or the amount of math professional development they attended (Table 6). High and low performers, however, sought out quite different colleagues for advice about math (Table 7). For example, high performers were less likely to seek out school leaders for advice than were low performers; high performers were also significantly less likely to seek out female staff members, colleagues with advanced degrees, and math coaches for advice about math, compared with their low-performing colleagues (Table 7).

Ruling Out Endogenous Sorting of Students Based on Prior Performance. Teacher performance, as measured using student test scores, includes not only the teacher's contribution to student achievement but also those students' prior achievement levels. If students were sorted to teachers based on their prior performance, our measures of performance would not only capture teacher performance but also the impacts of this sorting. To test whether students were systematically sorted to teachers in Auburn Park based on students' prior performance, we conducted a series of F tests to test whether there were significant differences between teachers within schools in the *incoming* math achievement of their students. We conducted these analyses for both district and state test scores, and separately for each school, controlling for whether a teacher

TABLE 7

Characteristics of Staff Sought Out for Math Advice by Low and High Performers, Auburn Park, 2010–2015

	Low performers (bottom quartile)	High Performers (top quartile)
Principal	0.27	0.13**
	(0.44)	(0.34)
Leadership role	0.77	0.56***
	(0.42)	(0.50)
Female	0.94	0.90*
	(0.23)	(0.31)
Advanced degree	0.67	0.59*
	(0.47)	(0.49)
Years experience in education	11.57	13.10†
	(12.60)	(11.07)
Years experience in school	7.08	7.54
	(12.74)	(9.89)
Multiple grades teacher	0.23	0.13***
	(0.42)	(0.34)
Math coach	0.07	0.03**
	(0.26)	(0.17)
n	571	516

Note. Analyses pool data from 2010, 2011, 2012, 2013, and 2015. High and low performers are determined separately for each year using the distribution of class average student math test scores on district tests. Tests of equality between columns account for repeated observations of some individuals.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

was a special education teacher, as these teachers might be expected to receive students with different prior achievement. We pooled together all 6 years of data for these analyses. As we conducted numerous tests, we corrected the critical values for these *F* tests to account for multiple comparisons.

For district test scores, results of these *F* tests were significant for one of the 14 elementary schools in Auburn Park; for state tests, the *F* test was significant for a different single school. The results of our analyses remained unchanged when these schools were excluded, suggesting that in Auburn Park, students were not sorted to classrooms based on students' prior achievement to a degree that significantly impacted our results.

Discussion and Conclusion

Our study makes several contributions to understanding teachers' thinking about their peers' expertise, the role of student test data in that thinking, and the degree to which colleagues' students' test scores predict interactions about instruction with peers. First, and consistent with prior work, our analysis suggests that perceptions of a peer's expertise influences whom school staff seek out for advice about instruction. Second, and extending prior research, our analysis identifies several ways in which school staff construct colleagues as expert math teachers, including direct and indirect knowledge of their mathematics instruction; their knowledge of, and disposition toward, mathematics; and holding a leadership position or formal training related to mathematics. Furthermore, our account captures how the school district's educational infrastructure enabled teachers to "see" and engage with these indicators of their peers' expertise.

A third and related contribution is our finding that, from the perspective of school staff, student test data is generally not regarded as a trustworthy measure of their peers' expertise. Our interviews showed that, despite the educational infrastructure that the district built to make student test scores available to staff, they did not rely on student test scores to a significant degree to identify the best mathematics teachers among

their peers; instead, other indicators of performance took precedence. Furthermore, while some school staff (less than half), when probed, did acknowledge using student test data as an indicator of teacher expertise, most reported doing so selectively, particularly to identify the weakest math teachers rather than the best. School staff had no difficulty in naming who they thought were the "best" mathematics teachers in their buildings, but felt that relying on student test data was not a good means of identifying these teachers, despite the significant supports that the district created that encouraged them to do so.

A fourth contribution is that teacher performance, as measured by their students' performance on state and district tests, does not predict being sought out for advice about mathematics teaching. Our finding here is inconsistent with findings from the sole prior study we found on the role of student test data in predicting teachers' instructional interactions (Wilhelm et al., 2016). There are several possible explanations for this, all relevant to interpreting and generalizing our findings. One is that our study focused on elementary schools rather than middle schools; another is that, unlike Wilhelm and colleagues' (2016) study, our study was undertaken in a district that was not under the pressure of high stakes accountability associated with NCLB. District and school leaders in our study, in contrast to the schools in Wilhelm et al.'s (2016) study, did not mention high stakes accountability in their discussions of the factors that influenced their interactions with colleagues about math. Several studies show that the pressure of high stakes accountability and related tests may influence interactions among school staff (Daly & Finnigan, 2011; Valli & Buese, 2007). Furthermore, the Wilhelm et al. (2016) study focuses solely on new, unreciprocated ties, while our study focuses on both new and existing ties. Another important difference is that in Auburn Park, the elaborated educational infrastructure—especially school organizational routines such as the grade-level PLC and team meeting routine, but also coaches and teacher leader positions—may have ensured that teachers not only interacted with teachers whose students performed at different levels but also provided opportunities for teachers to access and observe their colleagues' expertise in a variety of ways that went beyond student performance. Auburn Park teachers, for example, were required to interact with their grade-level peers weekly as part of the PLC and grade-level team meeting routines, regardless of their peers' performance levels. In this way, the district's educational infrastructure opened up opportunities for teachers to not only seek advice from peers whose students performed at varying levels, but also to observe firsthand other potential measures of their colleagues' instructional expertise, such as their passion for mathematics instruction and their accounts of their mathematics teaching. Our qualitative interview data offers evidence in support of this interpretation, showing how the school district's educational infrastructure influenced teachers' notions about the best math teachers in their school. A key takeaway from this finding, then, is that a school system's educational infrastructure, including its organizational routines and formal positions and training, do not just create venues in which teachers can learn from their peers, but also create venues in which they *can identify* expert teachers in a particular curricular domain.

Although some work suggests that higher performers are sought out less because they are less available to their colleagues as they are working on improving their own performance (see Baker-Doyle & Yoon, 2010), we found no evidence of this in our qualitative interview data, nor did our network data show that high performers gave less advice to colleagues than other teachers. To the contrary, in identifying the best mathematics teachers, school staff often referenced these teachers' availability and their willingness to give their time. Again, we suspect that the district's educational infrastructure, which created structured opportunities for staff to interact with one another through regular organizational routines, may have contributed to this availability. Furthermore, the fact that Auburn Park schools did not face intense high stakes accountability pressure tied to student test scores may have also contributed to the relative availability of high performers to their colleagues.

There are several possible reasons why teacher performance, as measured by student test scores, did not predict being sought out for

advice in Auburn Park. One reason, suggested by our interviews, is that teachers do not trust student test scores as valid measures of teacher performance in general and of the sort of mathematical learning pressed by Auburn Park's educational infrastructure in particular. Another possibility is that, compared with other ways of identifying teacher expertise, student test scores may not be as easily remembered or accessed as teachers make decisions about whom to seek for advice, especially if such decisions are made quickly during the work day. As suggested by our analysis of interview data, this is in part a function of an elaborated educational infrastructure that supported regular interactions among teachers about mathematics instruction. Yet another possibility is that knowledge of colleagues' test scores only impacts teachers' thinking about their colleagues' expertise over longer periods of time than captured in our study; as we generally use the 1-year lag of test scores to predict interactions, we may be missing these longer term impacts. These possible explanations are not mutually exclusive and potentially work together to account for why student performance did not predict being sought out for advice about instruction.

A fifth contribution of our work stems from our finding that teachers whose students performed better on state and district mathematics tests were more likely to *seek out* colleagues for advice and information about mathematics. This finding contradicts some prior work which suggests that more expert teachers seek *less* advice from colleagues, compared with less expert teachers (Bridwell-Mitchell & Cooc, 2016; Coburn et al., 2010). Our findings, in contrast, suggest that Auburn Park teachers whose students perform better, as calculated as the percentage of proficient students or by students' average test scores, positively predicted *seeking* advice about math in the following year. The better a teacher's students performed, the more likely she or he was to *seek* advice from colleagues the following year. One possible explanation for this is that high performers in general, and as measured in terms of their students' achievement in particular, represent a group of teachers who are constantly striving to improve by seeking out advice and information from others. Another possible explanation is that teachers whose students do well on district and state assessments assign their students more challenging work that may contribute to students' reacting less positively in class. Noticing students' less than positive reactions may prompt these teachers to seek out advice and information in an effort to "improve" their teaching. Although speculative, there is some evidence from university teaching to support this interpretation (Carrell & West, 2010).

In contrast to our other two test-based measures of teacher performance, teacher value-added did *not* predict advice seeking in Auburn Park. Because, as discussed earlier, value-added implicitly controls for student sorting to teachers, one possible explanation for this finding is that it was teachers who were assigned higher performing students, as opposed to those whose students performed better, who sought more advice. Our results, however, also show that there was not significant within-school sorting of students to teachers in Auburn Park based on students' prior achievement, suggesting that student sorting is unlikely to be an explanation of the differences between value-added and our other performance measures. Of course, it is also possible that our value-added measures were simply too noisy to capture any underlying differences in performance that were captured by our other performance measures.

Our account suggests that district policymakers can influence teachers' access to their expert peers, but relying solely on making student test data widely available to teachers is likely to be insufficient, due in part to the skepticism among school staff about test data as a metric for expertise. Instead, policymakers should attend carefully to the various ways that school staff use different sorts of information, including student test scores, to figure out instructional expertise among their peers, and design educational infrastructures that maximize opportunities for teachers to *see* and *engage* with these experts. As captured in the case of the district we studied, various components of a district's educational infrastructure—from organizational routines to formal leadership positions—can work together to create opportunities for school staff to see their peers' expertise and thereby influence who they seek out for instructional advice.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Work on this article was supported by the NebraskaMATH Study (see http://www.distributedleadership.org/projects.html) at Northwestern University and University of Nebraska–Lincoln, funded by research grants from the National Science Foundation (DUE-0831835). The work was also supported by the Distributed Leadership Studies (see http://www.distributedleadership.org) funded by research grants from the National Science Foundation (REC–9873583, RETA Grant No. EHR–0412510). Northwestern University's School of Education and Social Policy and Institute for Policy Research also supported this work. All opinions and conclusions expressed in this article are those of the authors and do not necessarily reflect the views of any funding agency.

Notes

1. As we only had student test scores for teachers of tested grades, we were unable to compare the associations between advice interactions and test scores for teachers of tested and nontested grades.

2. Personal communication with district official (M. M.), May 23, 2016.

3. Personal communication with district official (M. M.), January 15, 2015.

4. Personal communication with district official (D. R.), November 18, 2017.

5. A detailed discussion of these models is beyond the scope of this article; see Hoff et al. (2002) and Sweet et al. (2013) for further details.

6. Estimates from these models are also known to be relatively stable (up to certain identifiable class) and have desirable statistical properties based on the assumption of conditionally independent ties given latent positions.

7. In separate analyses not presented here, we divided the value-added distribution into quartiles and quintiles, and examined whether teacher value-added in the highest quartile (or quintile) was predictive of providing or seeking advice. We also examined whether the difference in performance between advice provider and seeker was predictive of a tie. In neither case did we find a pattern of significant associations. Results are available from the authors upon request.

References

Achinstein, B. (2002). Conflict and community: The micro politics of teacher collaboration. *Teachers College Record*, *104*, 421–455.

Adler, P. S., & Kwon, S. W. (2002). Social capital: Prospects for a new concept. *The Academy of Management Review*, *27*, 17–40.

Baker-Doyle, K. J., & Yoon, S. A. (2010). Making expertise transparent: Using technology to strengthen social networks in teacher professional development. In A. Daly (Ed.), *Social network theory and educational change* (pp. 115–126). Cambridge, MA: Harvard Education Press.

Becker, H. (1990). Generalizing from case studies. In E. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 233–242). New York, NY: Teachers College Press.

Blau, P. M. (1955). *The dynamics of bureaucracy: A study of interpersonal relations in two government agencies*. Chicago, IL: The University of Chicago Press.

Blau, P. M., & Scott, W. R. (1962). *Formal organizations: A comparative approach*. San Francisco, CA: Chandler Publishing.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2014). Ucinet. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of social network analysis and mining* (pp. 2261–2267). New York, NY: Springer.

Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). New York, NY: Greenwood.

Bridwell-Mitchell, E. N., & Cooc, N. (2016). The ties that bind: How social capital is forged and forfeited in teacher communities. *Educational Researcher*, *45*, 7–17.

Carrell, S. E., & West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, *118*, 409–432.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *The American Economic Review*, *104*, 2633–2679.

Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, *23*, 145–170.

Coburn, C. E., Choi, L., & Mata, W. (2010). "I would go to her because her mind is math": Network formation in the context of district-based mathematics reform. In A. J. Daly (Ed.), *Social network theory and educational change* (pp. 33–50). Cambridge, MA: Harvard Educational Press.

Coburn, C. E., Mata, W. S., & Choi, L. (2013). The embeddedness of teachers' social networks: Evidence from a study of mathematics reform. *Sociology of Education*, *86*, 311–342.

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

Cohen-Cole, E., & Fletcher, J. M. (2008). Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics*, *27*, 1382–1387.

Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, CA: SAGE.

Daly, A. J., & Finnigan, K. S. (2011). The ebb and flow of social network ties between district leaders under high-stakes accountability. *American Educational Research Journal*, *48*, 39–79.

Donaldson, M. L., & Papay, J. P. (2015). Teacher evaluation for accountability and development. In H. F. Ladd & M. E. Goertz (Eds.), *Handbook of research in education finance and policy* (pp. 174–193). New York, NY: Routledge.

Eisenhart, M. (2009). Generalization from qualitative inquiry. In K. Ercikan & W.-M. Roth (Eds.), *Generalizing from educational research: Beyond qualitative and quantitative polarization* (pp. 51–66). New York, NY: Routledge.

Eraut, M., & Hirsh, W. (2010). *The significance of workplace learning for individuals, groups, and organisations*. Oxford, UK: University of Oxford.

Feld, S. L. (1982). Social structural determinants of similarity among associates. *American Sociological Review*, *47*, 797–801.

Frank, K. A., & Zhao, Y. (2005). Subgroups as meso-level entities in the social organization of schools. In L. V. Hedges & B. Schneider (Eds.), *Social organization of schooling*. New York, NY: SAGE.

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, *44*, 96–104.

Grossman, P. L., Wineburg, S., & Woolworth, S. (2001). Toward a theory of teacher community. *Teachers College Record*, *103*, 942–1012.

Hargreaves, A. (2000). Mixed emotions: Teachers' perceptions of their interactions with students. *Teaching and Teacher Education*, *16*, 8811–8826.

Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How teacher evaluation methods matter for accountability: A comparative analysis of teacher effectiveness ratings by principals and teacher value-added measures. *American Educational Research Journal*, *51*, 73–112.

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Network and psychological effects in urban movement. In A. G. Cohn & D. M. Mark (Eds.), *International conference on spatial information theory* (pp. 475–490). Berlin, Germany: Springer-Verlag.

Horn, I. S. (2005). Learning on the job: A situated account of teacher learning in high school mathematics departments. *Cognition and Instruction*, *23*, 207–236.

Ibarra, H. (1992). Homophily and differential access in an advertising firm. *Administrative Science Quarterly*, *37*, 422–447.

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

Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics*, *6*, 801–825.

Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher*, *44*, 105–116.

Kossinets, G. (2006). Effects of missing data in social networks. *Social Networks*, *28*, 247–268.

Leana, C. R., & Pil, F. K. (2006). Social capital and organizational performance: Evidence from urban public schools. *Organization Science*, *17*, 353–366.

Leenders, R. T. A. J. (1996). Evolution of friendship and best friendship choices. *Journal of Mathematical Sociology*, *21*, 133–148.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: SAGE.

Little, J. W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers College Record*, *91*, 509–536.

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis*, *15*, 129–151.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*, 531–542.

Marsden, P. V. (1987). Core discussion networks of Americans. *American Sociological Review*, *52*, 122–131.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415–444.

Mollica, K. A., Gray, B., & Trevino, L. K. (2003). Racial homophily and its persistence in newcomers' social networks. *Organization Science*, *14*, 123–136.

Monge, P. R., & Contractor, N. (2003). *Theories of communication networks*. New York, NY: Oxford University Press.

Moolenaar, N. M., Karsten, S., Sleegers, P., & Daly, A. J. (2014). Linking social networks and trust at multiple levels: Examining Dutch elementary schools. In D. Van Maele, P. B. Forsyth, & M. Van Houtte (Eds.), *Trust and school life: The role of trust for learning, teaching, leading, and bridging* (pp. 207–228). Dordrecht, The Netherlands: Springer.

Moolenaar, N. M., Sleegers, P. J., & Daly, A. J. (2012). Teaming up: Linking collaboration networks, collective efficacy, and student achievement. *Teaching and Teacher Education*, *28*, 251–262.

Parise, L. M., & Spillane, J. P. (2010). Teacher learning and instructional change: How formal and on-the-job learning opportunities predict changes in elementary school teachers' instructional practice. *Elementary School Journal*, *110*, 323–346.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Thousand Oaks, CA: SAGE.

Penuel, W. R., Frank, K. A., & Krause, A. (2010). Between leaders and teachers: Using social network analysis to examine the effects of distributed leadership. In A. J. Daly (Ed.), *Social network theory and educational change* (pp. 159–178). Cambridge, MA: Harvard University Press.

Penuel, W. R., Riel, M., Korbak, C., & Means, B. (2004, April). *Investigation of a social capital approach to the adoption of reform practices*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

Penuel, W. R., Sun, M., Frank, K. A., & Gallagher, A. (2012). Using social network analysis to study how collegial interactions can augment teacher learning from external professional development. *American Journal of Education*, *119*, 103–136.

Pil, F. K., & Leana, C. R. (2009). Applying organizational research to public school reform: The effects of teacher human and social capital on student performance. *Academy of Management Journal*, *52*, 1101–1124.

Pitts, V., & Spillane, J. P. (2009). Using social network methods to study school leadership. *International Journal of Research and Method in Education*, *32*, 185–207.

Rivera, M. T., Soderstrom, S. B., & Uzzi, B. (2010). Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *Annual Review of Sociology*, *36*, 91–115.

Shalizi, C. R., & Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, *40*, 211–239.

Shrum, W., Cheek, N. H., & Hunter, S. M. (1988). Friendship in school: Gender and racial homophily. *Sociology of Education*, *61*, 227–239.

Small, M. (2009). *Unanticipated gains: Origins of network inequality in everyday life*. New York, NY: Oxford University Press.

Smylie, M. A. (1995). Teacher learning in the workplace: Implications for school reform. In T. R. Guskey & A. M. Huberman (Eds.), *Professional development in education: Paradigms and practices* (pp. 92–113). New York, NY: Teachers College Press.

Spillane, J. P., Hallett, T., & Diamond, J. B. (2003). Forms of capital and the construction of leadership: Instructional leadership in urban elementary schools. *Sociology of Education*, *76*, 1–17.

Spillane, J. P., & Hopkins, M. (2013). Organizing for instruction in education systems and school organizations: How the subject matters. *Journal of Curriculum Studies*, *45*, 721–747.

Spillane, J. P., Hopkins, M., & Sweet, T. (2015). Intra-and interschool interactions about instruction: Exploring the conditions for social capital development. *American Journal of Education*, *122*, 71–110.

Spillane, J. P., Hopkins, M., & Sweet, T. (2017). School district educational infrastructure and change at scale: Teacher peer interactions and their beliefs about mathematics instruction. *American Educational Research Journal*, *55*, 532–571. doi:10.3102/0002831217743928

Spillane, J. P., Kim, C. M., & Frank, K. A. (2012). Instructional advice and information seeking behavior in elementary schools: Exploring tie formation as a building block in social capital development. *American Educational Research Journal*, *49*, 1112–1145.

Spillane, J. P., Shirrell, M., & Hopkins, M. (2016). Designing and deploying a professional learning community (PLC) organizational routine: Bureaucratic and collegial arrangements in tandem. *Les dossiers des sciences de l'éducation*, *35*, 97–122.

Spillane, J. P., Shirrell, M., & Sweet, T. M. (2017). The elephant in the schoolhouse: The role of propinquity in school staff interactions about teaching. *Sociology of Education*, *90*, 149–171.

Steglich, C. E. G., Snijders, T. A. B., & Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, *40*, 329–393.

Steinberg, M. P., & Donaldson, M. L. (2016). The new educational accountability: Understanding the landscape of teacher evaluation in the post-NCLB era. *Education Finance and Policy*, *11*, 340–359.

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*, *15*, 250–267.

Sun, M., Loeb, S., & Grissom, J. A. (2017). Building teacher teams: Evidence of positive spillovers from more effective colleagues. *Educational Evaluation and Policy Analysis*, *39*, 104–125.

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

Sun, M., Wilhelm, A. G., Larson, C. J., & Frank, K. A. (2014). Exploring colleagues' professional influence on mathematics teachers' learning. *Teachers College Record*, *116*(6), 1–30.

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

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

Sweet, T., Thomas, A. C., & Junker, B. W. (2013). Hierarchical network models for education research hierarchical latent space models. *Journal of Educational and Behavioral Statistics*, *38*, 295–318.

Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, *44*, 519–558.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

Wilhelm, A. G., Chen, I. C., Smith, T. M., & Frank, K. A. (2016). Selecting expertise in context: Middle school mathematics teachers' selection of new sources of instructional advice. *American Educational Research Journal*, *53*, 456–491.

Yin, R. K. (2005). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: SAGE.

Authors

JAMES P. SPILLANE is the Spencer T. and Ann W. Olin Professor in Learning and Organizational Change at the School of Education and Social Policy at Northwestern University. Spillane has published extensively on issues of education policy, policy implementation, school reform, and school leadership. His work explores the policy implementation process at the state, district, school, and classroom levels, focusing on intergovernmental and policy-practice relations. He also studies organizational leadership and change. Recent projects include studies of relations between organizational infrastructure and work practices in schools and the socialization of new school principals. He is a member of the National Academy of Education.

MATTHEW SHIRRELL is an assistant professor of educational leadership and administration in the Graduate School of Education and Human Development at George Washington University. His research focuses on the impacts of educational systems on individual and organizational learning and improvement, and the interactions of policy, school working conditions, school leadership, and teacher retention.

SAMRACHANA ADHIKARI is currently a postdoctoral research fellow in the Department of Health Care Policy at Harvard Medical School. Her research interests are in Bayesian hierarchical models, social network analysis, and causal inference, for applications in education, social sciences, and health care policy.

Manuscript received July 4, 2017
First revision received January 19, 2018
Second revision received April 20, 2018
Accepted June 3, 2018