Easy to Dance To: Solving the Problems of Teacher Evaluation with Peer Assistance and Review

JENNIFER GOLDSTEIN
City University of New York, Baruch College

Scholars and practitioners have long criticized teacher evaluation as ineffective. Peer assistance and review (PAR) alters traditional teacher evaluation, as master teachers conduct summative as well as formative assessment of beginning teachers and veteran teachers in need of intervention. Relying on data from a longitudinal case study of one urban district, this article describes key components of teacher evaluation with PAR, in particular how it differs from teacher evaluation as typically conducted by principals. Findings are reported across six key factors: time, professional development, transparency, labor relations, decision making, and accountability. Notably, a substantially higher level of accountability appeared present with PAR than prior to program implementation. In contrast to popular opinion, this study provides an example of teachers willing and able to engage in the summative evaluation of their peers, a key component of professionalism and professionalization. Implementation challenges and areas for future research are addressed.

Interesting how we survived all these years [with teacher evaluation] by saying she is a cheerful, enthusiastic teacher. Works well with colleagues. Easy to dance to. They give her a 3. (PAR Consulting Teacher)

Introduction

Formal teacher evaluation is traditionally the leadership responsibility of principals. Research on teacher evaluation has rarely questioned the automatic purview of administrators in matters of quality control, in part because teachers and teacher unions have rarely been willing to take responsibility for the evaluation of peers and in part because administrators have been reticent to

relinquish a task seen as central to their leadership. This article questions the hierarchical and bureaucratic chain of command in education by presenting an empirical study of peer assistance and review (PAR), an alternate model of teacher evaluation used in some school districts over the last two decades in which lead teachers conduct personnel evaluations of other teachers. It asks, How is teacher evaluation with PAR different from traditional teacher evaluation by a principal? The study is concerned with the dynamics of the program and how this effort differs from traditional teacher evaluation practices. Six key factors emerged to distinguish PAR from more traditional practices:

- How much time are evaluators able to devote to the leadership task of evaluation?
- How, if at all, is evaluation linked to professional development efforts?[1]
- How transparent is the process of evaluation?
- What is the role of the teacher union in the process of evaluation?
- How confident are evaluators in the evaluative decisions that they make?
- Under this system of evaluation, how accountable are teachers for their performance?

The article builds on the literature that demonstrates the flaws of traditional teacher evaluation and posits that the more professional model of PAR shows promise. The data, drawn from one urban district implementing PAR, are presented in a unique way. For each of the six factors, the situation is first presented as a problem, based primarily on literature but also on interview data. The ways that PAR addresses the problem are then presented using data from the district studied.

A Professional Model of Teacher Evaluation

Teacher quality is key to student learning (NCTAF 1996), yet our ability to improve teacher quality is limited by an organizational structure imported

JENNIFER GOLDSTEIN is assistant professor at the School of Public Affairs of Baruch College/CUNY. Her areas of specialization are educational policy and leadership, and her work focuses on teacher workforce quality, teacher professionalization, and the distribution of leadership in schools and school districts. She previously worked as an elementary school teacher in Compton and Campbell, California. She recently published "Making Sense of Distributed Leadership: The Case of Peer Assistance and Review" in *Educational Evaluation and Policy Analysis* (2004) and "A Thoughtful Approach to Teacher Evaluation" in *Educational Leadership* (2006; with Pedro Noguera).

from the industrial factory model (Callahan 1962). Education became routinized and standardized, with layers of management for supervision, and administrator jurisdiction over teacher evaluation became the institutional norm (Lortie 1975; Tyack 1974). Feminist scholars have argued that teachers' work became organized in this hierarchical way for primarily paternalistic reasons (Grant and Murray 1999). The process that divided educational administration from teaching at the turn of the twentieth century has been considered a strategy that male teachers used to set themselves above their mostly female peers (Abbott 1988), resulting in an organizational structure that would have been far less likely for a predominantly male occupation (Grant and Murray 1999). Little (1988) claims that teachers, under this structure, have traditionally viewed professional obligations to one another as intrusive at worst and loosely invitational at best, what Feiman-Nemser and Floden (1986) term a "norm of noninterference." With this view, the responsibility for maintaining teacher quality and, hence, responsibility for evaluation reside hierarchically above teachers in the chain of command, in administration.

A countervailing vision for education, however, could alter this traditional hierarchy and center on the professionalism of teachers, vesting them with authority and responsibility for the quality of practice. In a more autonomous and professional vision of teaching, "teachers would in fact expect to be their brothers' keepers" (Little 1988, 94). Van Maanen and Barley (1984, 309) argue that "occupational self-control," or self-regulation, the autonomy held by an occupational community with respect to deciding "who will and will not be a member [gatekeeping], as well as how the content and conduct of a member's work will be assessed [quality control]," is central to a profession's development. Without this occupational self-regulation, teaching remains a "semi-profession" (Etzioni 1969).

In addition to occupational self-regulation, professionalism is generally characterized by a shared knowledge base and a concern for client welfare (Abbott 1988; Darling-Hammond 1990; Freidson 1986). A distinction can be made, however, between professionalism and professionalization. While professionalism is made up of traits possessed by the individuals of the profession, professionalization is the authority granted to the profession by society (Darling-Hammond 1997; Englund 1996). Self-regulation, or collective responsibility for professional standards, pertains to both professionalism and professionalization, however, as it involves a trait possessed by the individuals of the profession (i.e., the courage to self-regulate) as well as an external granting of authority that results in the formal ability to act on the possessed trait. Professionalism and professionalization are intimately linked, and Darling-Hammond (1997) warns that, without expertise in a shared knowledge base, professional autonomy will lead to bureaucratization rather than professionalization.

The Policy of PAR

PAR (pronounced as the word "par" and also referred to as "peer review") experienced a very specific birth in Toledo, Ohio, in 1981. Over the next two decades, a handful of districts—Cincinnati and Columbus, Ohio; Poway and Mt. Diablo, California; Rochester, New York; Dade County, Florida; and Salt Lake City, Utah—undertook the "Toledo Plan" of peer review, still the most well-known blueprint of the policy.[2]

Most commonly, PAR involves "consulting teachers," or CTs—teachers identified for excellence and released from full-time teaching duties for two to three years—who provide mentoring to teachers new to the district or the profession and intervention for identified veteran teachers experiencing difficulty. CTs also conduct the formal personnel evaluations of the teachers in the program. Teachers in either the new or veteran category are called "participating teachers," or PTs. The consulting teachers report to a district-wide joint teacher/administrator board, called the "PAR panel." The panel is typically cochaired by the union president and the director of human resources. Veteran teachers are most typically placed in PAR for intervention upon receiving an unsatisfactory evaluation from the principal, although in some districts other avenues for referral exist, such as a recommendation by the union building representative. Intervention cases are reviewed for validity by the PAR panel at the outset; the shortcomings in the teacher's performance must involve instructional matters, as noninstructional matters are not the purview of the PAR panel.

The panel holds hearings several times a year, at which CTs provide reports about PT progress. At the spring hearing, and sometimes sooner, the CTs make recommendations about the continued employment of each PT. A PT must meet specified quality standards within a set period of time, usually one year, or face removal from the classroom, as determined by the panel based on the recommendation of the consulting teacher, sometimes in concert with the principal. The panel's employment recommendation is passed to the superintendent, who makes a recommendation to the school board, the ultimate arbiter of personnel decisions.[3]

In 1999, California Assembly Bill IX (AB IX) marked the first time PAR was instituted statewide and the first time a major district had implemented the policy in over a decade. While California districts created PAR programs that looked quite different from one another, the district presented in this study matches the Toledo model.[4] By 2002, a state budget crisis and competing state legislation had begun to chip away significantly at California's PAR programs. This article is drawn from a study undertaken in that window of time, conducted in one urban district in California as it implemented the new PAR state policy. Involving a year of full-time fieldwork and data that span

four years, this study is among the most in-depth studies of a PAR program to date.

Study Design and Methods

The study used a single-case design (Yin 2003) of one urban school district in California, given the pseudonym Rosemont. Rosemont selected 10 consulting teachers (CTs) for the first year of the PAR program, who supported 88 beginning teachers and three veteran teachers across 28 schools. The primary study involved a year and a half of data collection, with follow-up data one year and three years later.

The sample included the district's nine members of the PAR panel (teachers and administrators) and 10 PAR CTs. In addition, three of the 10 CTs were chosen for more in-depth data collection. The study included PTs and principals based on their connection to the three case study CTs, as well as additional principals and PTs who might represent divergent or unrepresented viewpoints (Miles and Huberman 1994).[5]

The study relied on observations, interviews, and surveys. All panel meetings and hearings (approximately monthly) and almost every CT meeting (weekly) were attended for one year, totaling approximately 311 hours; these meetings were tape-recorded and scripted. Semistructured interviews were conducted with all panel members and CTs multiple times, as well as with 11 principals, 15 PTs, and three key district-level administrators, for a total of 67 interviews; all but three of these interviews were tape-recorded and transcribed.

Ongoing analysis of the qualitative data relied on summaries of field notes, analytic memoing, and coding (Miles and Huberman 1994). The software QSR NUD*IST 4 was used for data management. Coding schema were created from the researcher's progressive coding patterns. Multiple readers coded interviews against an initial thematic coding schema at the outset of analysis to verify reliability; once reliability was established, the researcher coded all subsequent interviews.

Finally, a multiwave survey approach complemented the continuous qualitative fieldwork during the first year of the program (Miles and Huberman 1994), with panel members, CTs, principals, and PTs surveyed. A survey was also collected from PTs at the end of the program's second year, providing longitudinal data and filling gaps left by the primary study. Only descriptive statistics are reported here (a methodological appendix is available in the online version of this article; for additional information about study design and methods, see Goldstein [2004]; and for the interview protocols and survey instruments, see Goldstein [2003]).

Findings

The examination of PAR in Rosemont yielded six key factors that distinguish it from teacher evaluation as typically conducted: (1) the amount of time spent on evaluation, where CTs support and evaluate a caseload of PTs full time; (2) the relationship that professional development has to evaluation, where evaluation is linked to support and professional development, including matching evaluators and evaluatees by grade and subject and using performance standards; (3) the transparency of the evaluation process, where PAR hearings and consulting teacher meetings make teachers' practice and evaluative decisions about that practice more transparent; (4) the nature of labor relations, where the teachers' union is part of the process; (5) the level of confidence in the decision-making process, where the PAR process seems to generate more confident evaluative decisions; and, ultimately, (6) the degree of accountability; while nearly everyone is retained with the current teacher evaluation system (Loup et al. 1996; Tucker 1997), PAR in Rosemont led to "nonrenewing" (firing) 12.5 percent of participating beginning teachers and 100 percent of the intervention cases in the first year of implementation.[6]

Throughout the article, survey responses for three groups of respondents—consulting teachers, panel members, and principals—are reported as one mean score on items rating PAR's effect on a variety of outcomes, as there were no significant between-groups differences. These survey responses involved Likert scales where 1 was a very negative effect and 5 was a very positive effect. The online version of this article contains appendices summarizing the survey ratings of PAR effectiveness.

Time

Problem.—Principals are phenomenally overwhelmed by the demands and expectations currently placed on them (Copland 2001; Grubb et al. 2003), with little time for instructional leadership precisely at a time when the focus on accountability for instructional results is very high. Noted one principal: "This is just the worst it's ever been in terms of my level of work. You feel so inadequate. You feel like, God, how come I can't do all this?!" This lack of time for instructional leadership limits the attention principals can bring to teacher evaluation (Painter 2000), and site administrators are often unable to find the time to thoroughly complete and document their teacher evaluations (Darling-Hammond 1984; Kelly 1998).

Principals admitted that they cut corners with their evaluations, by necessity. Principals described the "wiggle room" or need to be "creative" in doing their evaluations—typically doing fewer than desired or even required on teachers

perceived to be performing acceptably. One principal noted simply that "the current evaluation process really is a sham, it's a joke." Many principals identify their need to be in classrooms and know what is going on across the school but describe merely popping their heads in and out. Or, as this principal admits, some see teachers based on the whims of geography: "It probably depends how close they are to my office, too. Things as dumb as that even, whether they're on my trip. Like I'm going to go to the cafeteria in a few minutes and if they're on the way up, I'll probably see them more often than if they're over in the corner somewhere."

The lack of focus on teacher evaluation has clear implications for teaching and learning. Noted one teacher on the panel: "I was in a situation where I worked with a team member who was very unhealthy for the kids, and needed to be out of the classroom. [Other teachers and I] got to a point where we went to the principal and we said, 'You know this person is unsatisfactory. You know the kids aren't learning stuff. What's going on? Why didn't you give him an unsatisfactory [evaluation]?' 'I missed the deadline.' Or, 'Oh, I should have caught that. I'm just too busy.' Those are the kinds of excuses we got." Principals, as well as CTs and panel members, agreed that principals' lack of time allowed teachers not meeting standards to slip through the cracks with the traditional evaluation process in Rosemont.

Solution.—The CTs were released from full-time teaching responsibilities and focused on their PT caseloads. By contract, CT caseloads were 12–15 PTs. In reality, because CTs were involved in program development in the first year of implementation, they carried caseloads of approximately 10 PTs.[7] All CTs were expected to visit their PTs an average of one time per week, to make some unannounced visits, and to conduct three formal observation cycles during the year, presenting one at each panel hearing. PTs did report meeting with their consulting teachers on average once per week, especially at the start of the school year, but this ranged from "at least once a week" to once every two to three weeks, as consulting teacher visits to PTs' classrooms typically became less frequent for more effective PTs as the year progressed. Some CTs preferred to come by informally and unannounced, while others had a set time to visit every week. Noted one PT, "On Tuesday, we had a pretty routine schedule, which made it a lot nicer. I knew she was coming during 2nd and 3rd period every Tuesday, so I could count on that, I could make questions ahead of time that I knew I was going to want to ask. I'd teach during 2nd. So, she would typically observe during that time, and almost every time, she would give me written feedback on things that looked good and ideas for improvement. And then, 3rd period's my prep, so we could talk then." PTs reported that CTs made their ongoing accessibility clear at the beginning of the year, provided e-mail addresses and cell phone numbers, and could be reached as needed. Forty-seven percent of participating teachers and 80 per-

cent of principals interviewed initiated comments on the availability of the CTs and the amount of time they were able to spend working directly with PTs. The structure of CTs' full-time release from classroom teaching responsibilities allowed them to be on call to meet PT needs as they arose. Noted one CT, "There were a number of times where teachers called me on just specific little issues, whether it was a parent issue, a child abuse issue, an issue having to do with their principals, just little things, how tos, that were very simple to solve, but having that relationship was important."

Overall, CTs' time allowed a high level of involvement in the details of PTs' day-to-day lives that principals simply could not match, as they were busy running schools. A principal contrasted what she could provide to beginning teachers with what the CT provided: "Before PAR started I had Friday meetings with my new teachers and they would go forever, because they'd have a million questions and I would answer them and I would write down things that they needed and I would try to support them. But I can't model a lesson in every one of their classrooms and I can't do the kinds of things that a PAR consulting teacher can do because I'm running the whole school." The PTs recognized the difference between what their CT versus their principal could give them. Two of the 15 PTs interviewed had in fact had negative experiences with their principals and were therefore especially grateful to be involved in PAR. The majority of PTs, however, regarded their principals with respect for their seemingly insurmountable jobs and simply viewed the PAR program as a logical way for them to get desperately needed support. One PT made this compassionate contrast:

> My consulting teacher is a really good listener. I think more than my principal, my CT is a deeper, more thoughtful listener. She is doing something very specific for me, where my principal is doing a million things for everybody. And he always says, "Hey, whenever you want to come in, the door is always open, come in and talk to me." But you do get the feeling like "Hey, I'm really busy here, can we get this over with?" If my CT were principal it might be the same story. My principal wants to give me his attention, he's trying, I'm not going to say he's a total jerk who just sits there and says, "I don't have any time for you." I think he really wants to give his time. He's trying to do everything, but no one can do everything.

All stakeholders agreed that the structure of the PAR program, giving CTs full-time release from the classroom, allowed for a level of support to teachers that principals simply could not give.

Significantly, the link between comprehensive mentoring and new teacher retention is both commonsense and increasingly documented (Smith and Ingersoll 2004; Strong 2004). In addition, the combined group of principals,

panel members, and consulting teachers ($n = 34$) were asked to rate PAR's effect on principals' ability to do their jobs well, with a mean of 4.41 (SD = 0.56), where 5 was a very positive effect. Principals reported being able to relax a bit about their new teachers with the implementation of PAR, knowing the teachers were getting the consistent support and assessment they needed.

Support services are intimately dependent on the investment of time just described. A greater investment of time does not necessarily guarantee a trusting mentor-mentee relationship or high-quality guidance, but it may well be a necessary precursor.

The Relationship between Professional Development and Evaluation

Problem.—Teacher evaluation has generally been defined as a mechanism for appraisal in order to determine fitness for employment rather than a means for improving performance; the processes of formative and summative assessment are typically separated. Summative assessment occurs as a performance review, and formative assessment for professional growth may or may not occur depending on the setting. Where both occur, they are channeled through different leaders: an administrator is responsible for the summative review, while any number of support providers might be involved in professional development. Key here is that very often administrators conducting formal reviews are not privy to the knowledge and perspective of these support providers. Many authors have argued that formative and summative assessment are incompatible; they have raised concerns about the same person acting as both support provider and formal evaluator or about communication between people fulfilling these different roles (Costa and Garmston 1994; Nolan 1997; Popham 1988).

The dilemma is that principals' summative evaluations are therefore often based on very little data (Hunter 1988), limited to infrequent formal classroom observations that are almost always announced and may be quite short in duration. Compounding the problem, principals often lack specific content or grade-level expertise matched to their evaluatee (Darling-Hammond 1984), and they are often not well trained to conduct the evaluations (Loup et al. 1996; Wise et al. 1984). In addition, principals often use unclear or unstated performance criteria, and the criteria rarely reflect national or state curriculum or performance standards of best practice (Loup et al. 1996). Due in part to these issues, educators have complained that teacher evaluation is uneducative for teachers (Darling-Hammond 1984; Wise et al. 1984). As the CT in the epigraph wryly stated, "Works well with colleagues. Easy to dance to. They

give her a 3." At best, teacher evaluations are rarely focused on substantive development; at worst, they are meaningless. Noted one principal:

> So I come in and it's literally a Polaroid shot in time. I'm there for 50 minutes and I take my copious notes and stuff and I go back and now I meet with you and there's a form and it sort of says what were the barriers? What kept you from doing your lesson? Then what were some of the things that I saw? It's basically one way. It's sort of saying you're either doing an outstanding job or you're not cutting it and I'm giving you an A, B, C, highly satisfactory, blah, blah, blah, and tough luck.

With this system, evaluation is not about learning or developing as a professional but is merely the proverbial hoop through which to jump.

Solution.—As a result of CTs' full-time focus on PT support and formative assessment, summative assessments were based on ongoing observations throughout the year and intimate knowledge of a PT's classroom—rather than the notorious "dog and pony" show of most teacher evaluation systems. Linking evaluation to professional development through PAR (1) built trust and rapport, (2) provided PTs with ongoing instructional feedback, (3) created individualized support, and (4) grounded evaluation in performance standards for teaching and evaluator training for the CTs.

1. *Trust.* Most CTs felt that supporting PTs' day-to-day needs, especially at the beginning, helped develop rapport and build trust. While strong mentor programs often focus on trying to move mentor-mentee interaction beyond emotional support to substantive dialogue about teaching and learning, the reality remains that new teachers often do need emotional support (Gold 1996). For some PTs, the trust needed to speak openly about teaching and learning was developed by first knowing the CT was there to help. Noted one PT: "I think one benefit is just knowing that there is someone out there that is on your side, who you can go to to talk things through, to problem solve things."

In contrast to the traditional argument that formative and summative assessment are incompatible, linking support and evaluation did not appear to have a deleterious effect on PTs' trust in their CTs in most cases. Some authors have argued that these two leadership functions enhance one another (Hunter 1988; McGreal 1997; Stronge 1995). On the second-year survey, PTs' agreement that they trust their coach was a point and a half higher (on a four-point Likert scale) than their agreement that they cannot fully trust their coach. These findings were supported in PT interviews. While some reduction of trust did occur, it was limited to only a subset of particularly low-performing PTs (see Goldstein 2006a).

2. *Ongoing feedback.* In addition to building trust and rapport, however, the heart of the PAR program was ongoing feedback to PTs about how to teach.

Ninety-four percent of all interviewees (PTs, CTs, panel members, and principals) cited the feedback and suggestions on instructional strategies given by CTs to PTs, making it the most frequently named element of PAR in the research. Wherever possible, PAR CTs were paired with PTs by grade and subject matter. This is simply not possible with traditional teacher evaluation, as it would be rare for one principal to have the same curricular expertise as all of the teachers in his or her site. For several PTs, this matching was critical to their ability to work meaningfully with their CTs. Noted one: "The difference between my principal and [my PAR CT] is that my CT has experience in biology, and just in sciences in general, she was able to bring materials and suggestions to the class. The principal doesn't have that experience, her area isn't in sciences. My CT would make suggestions about how to go about teaching things, and it would trigger ideas and thoughts for me."

3. *Individualized support.* This grade and subject matching, together with the time CTs had available to work with PTs, created an environment of individualized support, which CTs often compared to a good teacher's ability to individualize instruction for students. The PTs noted that CTs had a high level of familiarity with day-to-day operations in their classrooms, allowing them to provide tailor-made support, such as bringing curricular materials that fit right in with a unit the PT was planning, being able to talk specifically about struggles with certain students, or recognizing when the PT was getting burned out and needed a break.

The individualized support provided to each PT contributed to informed evaluative judgments. CTs could look across a year from where a PT had started, what supports the PT had needed, what had been provided, and how the PT responded with what growth. In order to make these assessments, the work of the CTs was rooted in performance standards.

4. *Performance standards.* Strong evaluation systems include established standards for performance, rubrics, and evaluator training for interrater reliability (Tucker 1997). While CTs were not experts in performance standards for teaching at the time they were hired, they poured many professional development hours into becoming experts, and then into becoming calibrated among themselves in their use of the standards for evaluating to a rubric. PTs were evaluated on a slightly modified version of the California Standards for the Teaching Profession, which served as a benchmark throughout the year. Conversations between CTs and PTs about instruction were often grounded in standards language. Asked to rate PAR's effect on the use of teaching standards in the district, principals, panel members, and CTs ($n = 34$) had a mean of 4.60 (SD = 0.63), where 5 was a very positive effect.

Several principals were so impressed with the CTs' standards-based evaluations that they asked a CT to teach them the process. One CT described her expertise: "I really knew how to connect it all. I knew how to connect

the standards to the observations. I knew how to connect the observations to the individual learning plan [ILP]. I knew how to connect the ILP to the next observation. I knew how to connect the observation to the preliminary and the summary evaluation." Fluency in standards language gave CTs legitimacy with both principals and the panel, as well as with some PTs, and contributed greatly to the community's confidence in the quality of the evaluations being conducted.

Despite the efforts at calibration, it is possible that some PTs who were dismissed would have been retained had they had a different CT, and vice versa. In the first year of the program, as the CTs were working on calibration and protocols and they and the panel were beginning a discourse on the nature of quality teaching, variation undoubtedly existed between CTs. Nonetheless, as one principal noted: "I think there's a difference between any two people. What I like about PAR is that it does use standards. I'd have a lot of confidence that the difference between any two people evaluating one teacher is going to be really narrowed with PAR."

Taken together, the PAR program's components of support may in fact look quite similar to other full-time mentoring programs that do not involve summative evaluation. The study was not designed to look closely at interactions between CTs and PTs. Nor did the study compare PAR support to support in other mentor programs without evaluation. It cannot be said that the support provided by PAR is better than or even as good as other high-quality mentoring programs. For those interested specifically in mentoring, this is an area for further research. What can be drawn from this study, however, is that people involved in PAR believed that meaningful professional development was taking place as a form of evaluation that looked very different from teachers putting on a special show and principals flying through with a checklist.

Transparency

Problem.—Teaching has been an isolated occupation, with individual teachers behind closed doors with their particular group of students (Lortie 1975) and occupational norms that typically prevent teachers from "intruding" on one another's practice (Little 1988). Noted one principal, "The 11th Commandment is you don't speak ill of another teacher. I taught for seven years next to this nice person, just an awful teacher, and I could hear her through the wall, hear the kids and stuff and I would go over and have to quiet them down, just to kind of bring some sanity to it. But it was like the elephant in the living room. Nobody would talk about how awful she was."

In part because classroom practice has not been transparent, principals typically base their evaluations on minimal data. One principal highlighted

the situation this way: "Administrators, if you've got 35 people to evaluate, your contact is going to be limited to what's required. . . . Those times are going to be like snapshots. It's kind of like this: I'm going to come at 10:00 a.m. on March the 2nd to evaluate you in your room on a lesson that you choose to do with the class period that you choose. Will you be ready? You would think everyone would be ready!"

Next, alone with their observation notes or checklists, principals typically make evaluation decisions in isolation, not needing to defend their decisions to another colleague—let alone a panel of colleagues. In an ironic parallel to the lack of both support and oversight that plagues teaching, principals are neither held accountable for their evaluative decisions nor given the support that a group of colleagues can provide in the decision-making process.

Research has documented that principals give inflated ratings and few negative evaluations for a variety of reasons, including minimal observation data (Loup et al. 1996) and the desire to avoid conflict (Bridges 1986). One common practice is the "dance of the lemons," passing underperforming teachers to a different school (sometimes repeatedly for years) rather than giving a negative evaluation (Bridges 1992). Principals often know their teachers well; they may have worked together for years, and they may be friends. Especially considering the likelihood that a negative evaluation will not result in improved performance or a personnel change, principals' hesitancy to create conflict is understandable. It does little, however, to ensure a competent teacher for every student.

Solution.—PAR avoids some of the dilemmas of traditional teacher evaluation by first utilizing CTs across schools in the district based on grade and subject matching. In this way, CTs bring a broad, district-wide perspective to assessment, and a CT is not paired with a PT where there is a conflict of interest or other personal connection. Some smaller districts with PAR programs have formed consortia, pooling consulting teachers across districts in order to accomplish this goal.

Next, PAR opens the door to practice, since CTs are in teachers' classrooms on an ongoing and regular basis, generating more data on which to base the evaluations. Finally, PAR creates a formal team of colleagues and a structure for holding the evaluator accountable for his or her interpretation of the data.

1. *Opening the door to practice.* PAR alters the historic isolation of teaching by placing a mentor in PTs' classrooms on a frequent basis. While certainly not unique to PAR, the ongoing nature of PT-CT interaction is a critical piece in the quality of the evaluations, because increasing the publicness of practice is likely to increase the amount of information on which evaluations are based. CTs observed their PTs teaching across the year. Noted one PT:

> Had the vice principal come up to do the evaluation, she would have had no idea what it's like on a normal basis, when the vice principal

was not sitting in the back of the room. I really like the idea that my CT did my evaluations. Who better than someone who really has seen the whole picture? She had an idea of where I had started, and how much I had grown. She knew the struggles I had had, so she could look to see if I had addressed those. I really liked that there was some kind of benchmark.

Seeing PTs and their classrooms regularly, through both informal contact and formal evaluations including occasional unannounced visits, gave CTs the opportunity to see both growth and potential problems. One PT was relying on negative consequences for students, including listing the names of students along with their respective punishments on the board—which his CT discovered during an unannounced drop-in one day. The CT commented: "I think PAR strongly affects accountability. If [this PT] was an island, if he was on his own, doors closed, I never walked into that classroom, the names on the board would have continued for the rest of the year I guarantee it. He was required to meet the standard or at least show growth with the standard. It's the only reason why he changed some of the practices he was doing because that does not meet standards." PAR in Rosemont brought attention to teachers' day-to-day realities.

2. *Creating a team of colleagues.* Given a larger amount of data about a teacher on which to base both formative and summative assessment, PAR then creates a mechanism whereby multiple educators are in communication with one another about those assessments. CTs met as a group all day every Friday, and some of this time was spent discussing PT cases and seeking advice from one another. In addition, CTs formed pairs of "critical friends" and occasionally met to discuss their PT cases or visit a PT's classroom together for a second pair of eyes.

CTs also conferred with principals. CTs are focused on classroom practice, whereas principals have a perspective about the PT as part of the school community. By the second year of PAR, Rosemont created a format where both the CT and principal observed a PT and then conferred in order to be sure they were in agreement regarding professional development needed and/or the PT's assessment status.[8]

The most significant and formal public examinations of PT practice were the PAR panel hearings that occurred multiple times throughout the year. CTs reported, first with extensive documentation and then with oral presentations, on PTs' growth and/or problematic practice to the nine cross-district panel members approximately once each quarter. The panel offered suggestions of support that the CT might try and held the CT accountable for providing sufficient support in order for the PT to have the opportunity to improve. In this way, an individual teacher's practice became a district concern.

In a few instances, a CT was challenged to provide more evidence for a decision or even return to the PT for a few more weeks for one last effort. Noted one CT: "I was tap dancing around giving a decision of nonrenewal, and they asked me directly, 'What is the evidence for keeping this person?' And I really didn't have enough. They held me accountable, and that was appropriate."

One of the main criticisms of the PAR panel by CTs, however, was precisely that they did not play a critical enough role. For the most part, this seemed an issue of time. Hearings typically ran all day for two days. Yet most of those involved tended to feel the process was rushed, not allowing sufficient time to go into the depth they would have liked. It is therefore not surprising that some CTs reported feeling that the panel was a rubber stamp on their decision about a PT. While the study suggests that the transparency of the teacher evaluation process increased with PAR, there was still plenty of room for growth toward more meaningful involvement of the panel in the process.

The role of the panel is a very critical distinction between traditional teacher evaluation and PAR. We simply do not know how traditional teacher evaluation might be affected if principals had to defend their evaluative decisions to a board of colleagues or supervisors. The act of needing to defend one's decisions with evidence naturally alters and guides the decision-making process and can bring support to the evaluator. It is entirely conceivable that if principals were required to defend their evaluative decisions to a governing board and were given the time and training toward making those decisions, many of the changes documented here might also be seen. Similarly, CTs' evaluative decisions might look quite like principals' decisions without the oversight of the panel (Goldstein 2005).

Labor Relations

Problem. —The typically confrontational nature of education's labor relations makes the rare attempt at dismissal prohibitively costly and time consuming (Kerchner and Mitchell 1988; Kerchner et al. 1997; Sykes 1987; Urbanski 1999). Principals historically have viewed the union as an unbeatable adversary and often do not try to fire a teacher (Painter 2000). Instead, they engage in escape hatches (Bridges 1992), such as transfers (voluntary and involuntary), resignation, and retirement (Painter 2000). One principal explained that, with traditional teacher evaluation, "someone allowed me, not correctly, but allowed me to say you pick your battles and to be honest, you know, it's phenomenally hard to get rid of somebody. So I would say, 'Do I want to take the time to [get rid of them], knowing that I've also got this, I've got that, etc.' So you say, 'No.'"

Solution.—The AB IX legislation required that the union sign off on a district's proposal to the state creating a PAR program. In short, the teachers' union had the ability to prevent Rosemont from receiving state mentor money. The legislation also required that the panel be co-led by the union and district and made up of five teachers and four administrators. In these ways, the teachers' union played a central role in the changes brought about with PAR. Rating the effect of PAR on relations between the teachers' union and the district in Rosemont, the combined group of principals, panel members, and CTs ($n = 32$) had a mean of 4.44 (SD = 0.72), where 5 was a very positive effect. One principal highlighted the change with PAR:

> It takes forever to move teachers out if they're not doing the job. It's really hard as an administrator by yourself without PAR to do that. They grieve you and you have to do progressive discipline. All your eggs need to be in a row and the union really sort of comes at you. And this way [with PAR], I'm working collaboratively with the union. It's a whole different feel and there's a sense that the union and I agree that we need teachers who use best practice, and we're working together to have best practices occur, and we're not opposed in terms of keeping some person in there who is not utilizing best practice. I feel like we're all on the same team and it's about children and the kind of teaching they get.

Some principals were quite surprised to see the teachers' union president sitting at the table at hearings, let alone arguing for dismissals of teachers. PAR programs, however, have historically been initiated by union presidents interested in "postindustrial unionism" (Kerchner et al. 1997), and it was the union president who advocated for the creation of a PAR program in Rosemont prior to the implementation of AB IX. For some teacher unions, PAR is one way to defend the profession of teaching rather than individual teachers (Kerchner et al. 1997).

Decision Making

Problem.—Principals often doubt themselves when making evaluative decisions (Bridges 1986). How could it be otherwise? The problem of making a decision has accrued through the prior problems. Principals do not have sufficient time to spend on evaluations and are not involved in professional development in an ongoing and substantive manner; therefore, they are uncertain that the teacher under review has been given an opportunity to improve. They typically lack standards on which to rate teachers. They are alone to make the decision, without the benefit of an organizational structure that

provides collaboration with colleagues. Finally, they know that a negative evaluation will involve a timely and costly battle with the teacher union and that they will likely lose that battle.

One extreme case involving a middle school PT whose performance was clearly below standards exemplified this phenomenon. The PT was nonrenewed, and in the appeals hearing before the PAR panel in which he was entitled to present his case he was asked what different methods he would employ if he were to return to his classroom. He responded, "I would take over the classroom right away—that it was my world—and that if [the students] were entering it, they would be following my rules. I would also yell and scream a lot more, it seems to work very well." His principal was relieved by the outcome of his nonrenewal, yet revealed in her interview that even in this case she most likely would not have been confident enough to make the decision to nonrenew: "I needed [the PT] to be gone. I think it was easier for that to happen because he was in PAR. Without PAR, he might have slid through and be back next year, because I wouldn't have been able to give him the one on one all the time that the CT did. It would have taken a lot of time to identify all the problems and do all the documentation. And as a new principal, doubting myself, because he's a first year teacher, maybe I didn't do enough. And he might have grieved. That's always a legal weight." This principal acknowledged that doubt surrounds the accuracy and defendability of what might be considered the clearest of evaluations.

Solution.—Just as the problem of making a decision accrues through the prior problems, so the solution accrues through the prior solutions. Owing at least in part to the amount of time devoted to supporting the PTs, the ongoing nature of the evaluations, the link between the evaluations and teaching standards, and the shift from one evaluator standing alone to a group of peers participating in the process, CTs, principals, and the panel had an increased sense of confidence in the quality and accuracy of the evaluations themselves. Perhaps one of the most significant findings in the study is that, across the board, CTs were seen to be willing to recommend nonrenewal of a PT. This is not to imply that CTs were eager to recommend nonrenewal or that they did not agonize about such decisions when they had to be made. Certainly the role of evaluator is fraught with tension. Nonetheless, CTs rose to the challenge—not in all cases but at a much higher rate than principals. Rating PAR's effect on teacher evaluation in the district, the combined group of principals, panel members, and CTs ($n = 34$) had a mean of 4.60 (SD = 0.70), where 5 was a very positive effect. While the study did not examine the teacher evaluation paperwork, people involved in PAR, including principals, believed that higher-quality evaluations were being conducted than had traditionally occurred.

Teacher Accountability

Problem.—Teachers are rarely fired for teaching poorly (Loup et al. 1996; Tucker 1997). In one study, merely 0.1 percent of teachers were dismissed, despite the fact that 1.53 to 2.65 percent were formally identified as "incompetent" and 5 percent were informally identified as "incompetent" (Tucker 1997). Teachers are more likely reassigned to other school sites (Bridges 1992; Tucker 1997).

Solution.—CTs were recommending nonrenewal, principals and panel members had confidence in their recommendations, and the teachers' union was part of the process rather than against it. The result was that out of 88 new teachers who were in the program, eleven (12.5 percent) were nonrenewed for employment in year 1 of the program, a stark contrast to the automatic granting of tenure that often meets new teachers after a set number of years of service (Bridges 1992; Peterson 1995). In addition, three out of three veterans (100 percent) were encouraged into retirement.[9] This constituted a major change in accountability when compared to prior dismissal rates in the district, supporting findings in prior PAR studies (Darling-Hammond 1984; Kelly 1998; Murray 1999). In the year immediately before PAR, only three teachers out of a teaching force of almost 3,000 (0.1 percent) were nonrenewed, a figure identical to that reported in the Tucker (1997) study. While some teachers were removed for noninstructional reasons, such as tardiness or drug problems, the union president could not recall (and the district had no record of) any teachers being dismissed for issues of teaching quality in the years prior to PAR. The union president, calling one new teacher's dismissal "historic," noted:

> The fact is that for the first time people are actually being let go for mediocre teaching performance. This is the first time in this district that that's happened, that someone has been let go from the district for just winging it, not really being a clear and present danger to children, but not really getting it about what the job entails. Not because they were evil people or molested children or anything like that. Not because someone took a personality dislike to them, not because of this, that, or the other thing. Simply because their classroom performance was mediocre.

The full import of the higher dismissal rates is seen in the context of a year of attempted support. Far from a draconian or capricious decision, a PAR dismissal represented a concerted and collaborative effort to help a teacher improve that ended with a decision that the teacher's improvement was beyond the ability of the district. CTs and panel members often noted that they were

fulfilling a responsibility to the students of the district, in effect "stepping up" to do a difficult job that had to be done. Noted one teacher on the panel, "I left the [April] panel hearing feeling like, wow, we've done the district a favor in removing new teachers that could have stayed on, just taking up a place and becoming tenured. I just feel like we've really done a good job cleaning house this year. I'm excited about it."

Summary of Findings

Table 1 summarizes the differences between traditional teacher evaluation by a principal, as reported in the literature and the participants in this study, and evaluation through PAR in Rosemont. This was certainly not a comparative study; the intent here is merely to revisit in summary the key findings of the study alongside conventional wisdom regarding traditional teacher evaluation. It is worth reiterating that if principals were given the same time to conduct evaluations, the same training, and the same collaborative support as the CTs in Rosemont, teacher evaluation by a principal would most likely look very different. In other words, it would be a gross simplification to claim that master teachers are better able than principals to conduct teacher evaluations. Rather, the PAR program presented here addressed structural barriers in the system of teacher evaluation that allowed the CTs to achieve results that principals are typically unable to achieve.

In addition, PAR programs look different in different locales. Paper-driven "PAR" programs exist that look much like traditional teacher evaluation, simply substituting the administrator with a consulting teacher. Conversely, some so-called PAR programs involve CTs in only formative, not summative, assessment. The findings reported here are based on one particular PAR program and speak to PAR's promise; they are by no means universal.

Those involved in PAR in Rosemont felt confident enough in the soundness of the evaluative decisions to make them and stand behind them—despite being steeped in a culture where teachers do not speak ill of one another and nonrenewal decisions rarely take place. It is important, however, to note the difference between confidence in evaluations, or perceived quality, and the actual quality of the evaluations in this discussion. Given the level of analysis, the data from this study can speak only to the perceived quality of the PAR evaluation process, and not to the appropriateness (i.e., the reliability and validity) of the evaluations conducted or employment decisions made. It is possible that PAR provided the mechanism for strong rationalizations about personnel decisions that may not have been sound. However, PAR's data-driven rationalizations, completely sound or not in year 1 of implementation,

TABLE 1

Summary of Literature and Findings

	Traditional Teacher Evaluation	PAR in Rosemont
Time	Low Principals lack sufficient time to conduct evaluations or complete documentation (Copland 2001; Darling-Hammond 1984; Grubb et al. 2003; Kelly 1998)	High CTs support and evaluate a caseload of PTs full time
Professional development	Separate from evaluation Evaluation is separate from support and professional development and many principals are not well-trained to conduct educative evaluations (Darling-Hammond 1984; Hunter 1988; Loup et al. 1996; Wise et al. 1984)	Linked to evaluation Evaluation is linked to support and professional development, with grade/subject matching and the use of performance standards
Transparency	Isolated Teachers teach in isolation (Lortie 1975) and principals evaluate in isolation (Bridges 1986, 1992)	Public PAR hearings and CT meetings bring more eyes to teachers' practice and evaluators' assessments
Labor relations	Adversarial Confrontational labor relations makes dismissal prohibitively costly and time consuming (Kaboolian and Sutherland 2005; Urbanski 1999)	Collaborative The teachers' union is part of the process
Decision making	Hesitant Principals sidestep poor performance to avoid conflict (Bridges 1986) and doubt themselves	Confident CTs, principals, and panel members were mostly confident in the quality and accuracy of the evaluations, and CTs appeared willing to recommend dismissal
Accountability	Low Nearly everyone is found to be competent (Tucker 1997)	High 12.5% (11/88) of beginning PTs and 100% (3/3) of veterans were nonrenewed for employment in Rosemont

would nonetheless seem an improvement over principals' traditional rationalizations that are not based on evaluative data.

Discussion and Implications

Practitioners interested in implementing PAR, and researchers interested in studying it further, should carefully consider a number of challenges. This section briefly highlights seven PAR challenges for consideration.

Ensuring Consulting Teacher Quality

The perceived success of the program appeared to be based largely on principals' and panel members' confidence in the CTs. It follows that CTs should therefore be selected very carefully. CTs must be regarded as master teachers, and in Rosemont the selection process included classroom observations by two panel members. The CTs were also required to demonstrate prior success mentoring a peer, including a letter of recommendation from a teacher whom they had mentored. Finally, the CTs had to be above reproach. Given the authority that CTs held with respect to employment recommendations, it was critical that the selection process appear unbiased and without favoritism. Once selected, it was imperative that CTs received training in coaching methods, standards, and assessment and that they remained vigilant with respect to confidentiality.

Defining Good Teaching

Effective PAR programs require agreed-upon standards of practice and performance rubrics, which form the foundation of the work between participating and consulting teachers. In addition, evaluative decisions must be beyond reproach, which is supported by standards-based documentation for evaluations. The challenge in many districts is that educators have not defined quality teaching or made their implicit priorities and values explicitly clear—a necessary step for a transparent evaluation process. They also may not find themselves in agreement when they do make their values explicit. Creating these conversations, and owning rather than importing the standards of practice that grow out of them, is a crucial step in the PAR process.

Reframing Labor Relations

While both the American Federation of Teachers and the National Education Association officially support PAR, union locals are often not aligned to national leadership, and many teachers characterize PAR evaluators as "teacher cops." A critical issue for PAR implementation is the level of trust between teachers and administrators (Urbanski 1999). For this reason, most school districts begin PAR programs with new teachers only, since the idea of apprenticeship is far less controversial among teachers than peer evaluation for veterans. The expansion to include intervention cases typically occurs once a program has been successfully in place for a few years. This was not the case in California, where the state legislation specifically targeted veteran teachers. As a result, Rosemont and other districts across the state were required to skip the trust-building phase of PAR.

Rosemont's experience may serve as caution to other union presidents and district superintendents interested in forging new relationships between so-called labor and management. PAR is a great risk for union presidents, who are voted in and out by their membership. Respondents reported an improvement in labor relations with PAR, but the Rosemont union president lost an election a few years after the implementation of the policy—voting, after all, was not limited to those with firsthand knowledge of the PAR program. Some teachers characterized the president as too soft on management, and indeed the replacement president was a more traditional unionist who took a more adversarial approach. In this light, superintendents interested in PAR should recognize the political environment faced by union presidents and work to grant concessions in contract negotiations that union presidents can sell to their members, offsetting potential negative reactions by some teachers—in particular veterans—to the policy.

Reframing Instructional Leadership

Despite their complaints that they do not have time to do evaluations well, administrators are often quick to defend their turf. PAR potentially creates a tension for principals vis-à-vis their instructional leadership. The transition from authoritarian to participative leadership is a difficult one for principals (Kerchner and Koppich 1993), who are charged with instructional leadership and then asked to move over for leadership by teachers (Little 1988). Principals' hesitancy to relinquish authority for teacher evaluation is understandable and, where it signals professional commitment to teacher quality and instructional leadership, laudable.

The problem and its solution lie in the conception of instructional lead-

ership. Rather than define an instructional leader as one who directly provides the instructional support for teaching and learning, a distributed perspective on leadership (Heller and Firestone 1995; Leithwood and Jantzi 2000; Ogawa and Bossert 1995; Spillane 2006; Spillane et al. 2001) suggests that an effective instructional leader can generate the conditions for a focus on instructional matters, actively involving the leadership of those around her. With this framing, principals with PAR enact instructional leadership by communicating regularly with CTs, meeting with the panel, and conducting the personnel evaluations of those teachers not in PAR.

Building Bridges to Mentoring Programs

As outlined earlier, many educators adhere to the notion that formative and summative assessment must be separate in order to ensure trust between mentor and mentee. While that concern was not supported by this research, those interested in PAR must attend to it or face certain resistance. This resistance can be particularly strong among those focused on mentoring, who often view summative assessment as harsh treatment for new teachers. In California, opposition to PAR from within the state mentoring program played a role in undercutting the PAR program. Senate Bill 2042 introduced sweeping changes to teacher induction; the bill included a clause that prohibited any of the formative assessments from being used for summative evaluation purposes.

As already noted, PAR programs exist that do not include the focus on professional development reported here. Rosemont's PAR program benefited greatly by resting on a decade and a half of mentoring efforts in California, in particular by the Santa Cruz New Teacher Project and the statewide Beginning Teacher Support and Assessment (BTSA) program. Rosemont's CTs were able to enter an already existing statewide conversation about performance standards for teaching and effective coaching strategies, and some of them had already served as BTSA mentors. Strong PAR programs require deep knowledge about teaching and learning that is very often housed among those who may be PAR's biggest adversaries.

Paying for PAR

A comparison of PAR to traditional teacher evaluation should ideally address the cost differential. The main cost involved with PAR is the replacement cost of CTs who leave the classroom, which in Rosemont was covered by funding from the state per the state legislation. Other, more minor costs include stipends

for teachers on the PAR panel and release days for PTs to observe other teachers. These costs must be compared to the costs for traditional teacher evaluation. Officials in Rochester, New York, for example, estimated that with regular teacher evaluation they spend one-half to one full day of principal time for each probationary teacher each year, as well as an additional half day of clerical time (T. Gillett, personal communication, November 2004). There is a corollary, albeit lower, figure for tenured teachers. One must factor in any current expenses for induction and mentoring programs that would be terminated with PAR or folded into the PAR program. Finally, the legal costs for removing an unsatisfactory veteran teacher are $50,000–$200,000, depending on the state (Kaboolian and Sutherland 2005), and it usually takes three to six years for the litigation to run its course. Kaboolian and Sutherland (2005) report that effective peer review programs reduce litigation costs associated with terminating tenured teachers but highlight the complications for a comprehensive cost comparison. They note that peer review programs weed out weak teachers while they are probationary, avoiding the expense of termination later after they become tenured; peer review programs also improve retention, avoiding the expense of recruiting, hiring, and orienting yet more new teachers. These cost savings are hard to measure (Kaboolian and Sutherland 2005, 34); future research should nonetheless attempt to do so.

Overcoming the Norms of 100 Years of Bureaucracy

Despite the largely positive response to PAR in Rosemont, it is very difficult to shift norms in the way required by this policy. Despite the confidence in CTs' abilities and the perception of high-quality evaluations, most people— principals, panel members, and CTs themselves—wanted principals to be more involved in the process. For a discussion of this phenomenon, please see a companion piece (Goldstein 2004). Suffice it to say that policy makers and practitioners should be clear about their intentions regarding instructional leadership and responsibility for teacher evaluation when implementing PAR, as people will tend to regress to that which is familiar, namely, principal control.

Conclusion: Reframing Evaluation

This study of PAR demonstrates that teachers can and do evaluate one another. The transition to being one's brother's keeper is not easy (Kerchner and Koppich 1993; Kerchner et al. 1997; Wasley 1991). Nonetheless, CTs conducted summative evaluations and made decisions about the continued employment of other teachers. They reported those decisions to the panel, which

was responsible for the final decision to be recommended to the superinten-dent. However, across 91 PTs, the panel upheld CTs' employment recom-mendations in all cases.[10]

Yet, the role of CT is different from that of resource specialist or mentor teacher or other professionalization roles that officially elevate teachers into expert status. As was argued in the introduction, the gatekeeper function— holding responsibility for decisions about the quality of performance of others in one's profession—is key to being a professional (Darling-Hammond 1990; Van Maanen and Barley 1984). As such, PAR differs from other teacher leadership policies because it transcends the role expansion of the individual teacher leader to signal role expansion for the profession of teaching. Smylie et al. (2002, 163) note that current models of distributed leadership "depart from the individual empowerment, role-based models of teacher leadership that dominated the 1980s and early 1990s. They reframe teacher leadership as a more collective, task-oriented, and organizational enterprise."

Recall from the introduction that Darling-Hammond (1997) expressed con-cern that professional autonomy without expertise risks bureaucratization. With PAR in Rosemont, the data show a move to professionalization connected to demonstrated professionalism, as the district appears to have granted au-thority for evaluations to CTs based on perceived CT expertise; principals and panel members came to view CTs as very qualified to conduct teacher evaluations. CTs, having demonstrated their knowledge, expertise, and com-mitment to self-regulation (Darling-Hammond 1997; Sykes 1987)—in other words, their professionalism—were granted a degree of authority critical to professionalization.

In addition to a shared knowledge base and collective responsibility for professional standards, the third element of professionalism identified in the introduction was concern for client welfare. The CTs and panel members defined their function as improving the quality of teaching for the clients of the district—students. They expressed both a belief that PTs could improve and their commitment to helping them do so. If a PT's performance was ultimately not meeting standards, however, they saw their job as recom-mending dismissal of the teacher. While firing someone from employment is extremely difficult, CTs mollified themselves with the reminder of the greater good of improving teacher quality for students. It was the professional identity of those involved that enabled them to put client welfare, or at least their perception of it, at the forefront of their work. For Rosemont's CTs in par-ticular, for whom the renewal or nonrenewal decision weighed heaviest, con-cern for client welfare prevented taking the path of least resistance—continued PT employment—and therefore sometimes came before loyalty to a fellow teacher with whom they had worked for a year.

The emphasis on the firing of new teachers as "good news" may seem at

best cold-hearted or at worst irresponsible at a time when improving teacher retention is critical to improving teacher quality in urban schools (see Lankford et al. 2002). In a professional model of evaluation that includes a serious concern for client welfare, however, the goal cannot simply be retention, where we strive to keep in teaching anyone with a pulse. The goal is differentiated retention, where we strive to retain high-quality teachers (or those who show the potential to grow into high-quality teachers) and seek to remove from classrooms those teachers who are not performing up to standards and who show little promise of doing so. New teachers are more likely to stay in both teaching and their current settings if they are provided with the support they need (Smith and Ingersoll 2004), and the data presented here suggest that PAR may provide that support. It is nonetheless also important to take the responsibility for quality control and gatekeeping seriously, lest we continue to deprofessionalize teaching and fail to adequately serve children.

Self-regulation, central to professionalism and professionalization, has been slow to occur in education. Policy makers, practicing educators, and the public tend not to believe that teachers are capable of regulating themselves. Yet, with PAR in Rosemont, teachers' work expanded to include quality control and gatekeeping, with many benefits to the quality of the evaluation process— and potentially to the quality of teaching and learning.

Notes

1. Educational administration literature distinguishes between "evaluation" and "supervision" (Hazi 1994; Iwanicki 1998), although this literature lacks agreement as to whether supervision is properly an administrative role (McGreal 1997; Nolan 1997). While the most commonly invoked purpose of evaluation is the need to ensure a quality standard (summative assessment), various authors (Costa et. al. 1998; Danielson and McGreal 2000; Wise et. al. 1984) have encouraged educators to move beyond this narrow definition of evaluation to include the supervisory aspects of professional growth (formative assessment).

2. See peer review case studies of Toledo (Darling-Hammond 1984; Gallagher et al. 1993), Poway (Gallagher et al. 1993), and Rochester (Grant and Murray 1999; Koppich and Kerchner 1999; Murray 1999).

3. Many PAR programs also include an alternative evaluation option for tenured teachers who are meeting standards. The PAR program presented here, however, did not include this component at the time of the study.

4. For example, many California districts did not include new teachers in their PAR programs, as the state law required the program only for veteran teachers who had received an unsatisfactory evaluation from their administrator. Many programs did not create full-time positions for CTs, and many did not involve CTs in summative evaluation. Program details were left entirely to each district.

5. A note on the principals involved in the research: in the first year of implementation, district leadership placed PAR in schools where the principal had signed up for it.

6. More precisely, the three intervention cases chose to retire prior to being dismissed.

7. CTs developed a formula whereby supporting an intervention case counted as two beginning teachers when constructing caseloads, given what they perceived as the larger emotional drain and investment of time needed when working with a veteran teacher.

8. For an organizational analysis of Rosemont's division of responsibility for teacher evaluation between consulting teachers and principals, the changes in authority relations it entailed, and the eventual shift to a more formal role for principals in PAR evaluation, see Goldstein (2004).

9. The veterans placed in the program in its initial year were perceived to be notoriously below standards, the ones "we've been trying to get rid of for years." By the third year of the program, Rosemont successfully remediated one of the four veterans in PAR that year. This still placed the district below the average of a sample of other established PAR programs, where 30–60 percent of veterans were remediated (Darling-Hammond 1984; Hewitt 2000; Kelly 1998; Murray 1999).

10. This is accurate technically and substantively. However, an analysis of transcripts from panel hearings conducted for a companion study revealed that, in four cases, CTs altered their written recommendations to match their oral presentations to the panel, following challenges from panel members; see Goldstein (2006b).

References

Abbott, Andrew. 1988. *The System of Professions: An Essay on the Division of Expert Labor.* Chicago: University of Chicago Press.

Bridges, Edwin M. 1986. *The Incompetent Teacher: The Challenge and the Response.* Philadelphia: Falmer.

Bridges, Edwin M. 1992. *The Incompetent Teacher: Managerial Responses.* Washington, DC: Falmer.

Callahan, Raymond E. 1962. *Education and the Cult of Efficiency.* Chicago: University of Chicago Press.

Copland, Michael A. 2001. "The Myth of the Superprincipal." *Phi Delta Kappan* 82 (7): 528–33.

Costa, Arthur L., Robert J. Garmston, and Linda Lambert. 1998. "Evaluation of Teaching: The Cognitive Development View." In *Teacher Evaluation: Six Prescriptions for Success*, ed. Sarah J. Stanley and W. James Popham. Alexandria, VA: ASCD.

Danielson, Charlotte, and Thomas L. McGreal. 2000. *Teacher Evaluation to Enhance Professional Practice.* Princeton, NJ: Educational Testing Service.

Darling-Hammond, Linda. 1984. "The Toledo (Ohio) Public School Intern and Intervention Programs." In *Case Studies for Teacher Evaluation: A Study of Effective Practices*, ed. Arthur E. Wise, Linda Darling-Hammond, Milbrey W. McLaughlin, and Harriet T. Bernstein. Santa Monica, CA: RAND.

Darling-Hammond, Linda. 1990. "Teacher Professionalism: Why and How?" In *Schools as Collaborative Cultures: Creating the Future Now*, ed. Ann Lieberman. Bristol, PA: Falmer.

Darling-Hammond, Linda. 1997. *The Right to Learn.* San Francisco: Jossey-Bass.

Englund, Tomas. 1996. "Are Professional Teachers a Good Thing?" In *Teachers' Professional Lives*, ed. Ivor F. Goodson and Andy Hargreaves. Washington, DC: Falmer.

Etzioni, Amitai. 1969. *The Semi-professions and Their Organization: Teachers, Nurses, Social Workers*. New York: Free Press.

Feiman-Nemser, Sharon, and Robert E. Floden. 1986. "The Cultures of Teaching." In *Handbook of Research on Teaching*, 3rd ed., ed. M. C. Wittrock. New York: Macmillan.

Freidson, Elliot. 1986. *Professional Powers: A Study of the Institutionalization of Formal Knowledge*. Chicago: University of Chicago Press.

Gallagher, James J., Perry Lanier, and Charles Taylor Kerchner. 1993. "Toledo and Poway: Practicing Peer Review." In *A Union of Professionals: Labor Relations and Educational Reform*, ed. Charles Taylor Kerchner and Julia E. Koppich. New York: Teachers College Press.

Gold, Yvonne. 1996. "Beginning Teacher Support: Attrition, Mentoring, and Induction." In *Handbook of Research on Teacher Education*, 2nd ed., ed. John Sikula. New York: Macmillan.

Goldstein, Jennifer. 2003. "Teachers at the Professional Threshold: Distributing Leadership Responsibility for Teacher Evaluation." PhD diss., Stanford University, School of Education; Dissertation Abstracts International 64 (5): 1475A.

Goldstein, Jennifer. 2004. "Making Sense of Distributed Leadership: The Case of Peer Assistance and Review." *Educational Evaluation and Policy Analysis* 26 (2): 173–97; reprinted with apology; published originally in *Educational Evaluation and Policy Analysis* 25 (4): 397–421.

Goldstein, Jennifer. 2005. "Growing the Distributed Leadership Framework: 'Accoun-Table Distribution' for Teacher Evaluation." Paper presented at the annual meeting of the American Educational Research Association, Montreal, April.

Goldstein, Jennifer. 2006a. "Combining Supervision and Evaluation and Living to Tell about It: Mentee Trust and the Fear of Peer Review." *Journal of Personnel Evaluation in Education* 18 (4), http://dx.doi.org/10.1007/s11092-006-9022-3.

Goldstein, Jennifer. 2006b. "Designing Transparent Teacher Evaluation: The Role of Oversight Panels for Professional Accountability." Paper presented at the annual meeting of the University Council for Educational Administration, San Antonio, TX, November.

Grant, Gerald, and Christine E. Murray. 1999. *Teaching in America: The Slow Revolution*. Cambridge, MA: Harvard University Press.

Grubb, W. Norton, Joseph Flessa, Linda Tredway, and Jane Stern. 2003. "A Job Too Big for One: Multiple Principals and Other Approaches to School Leadership." Paper presented at the annual meeting of the American Educational Research Association, Chicago, April.

Hazi, Helen M. 1994. "The Teacher Evaluation-Supervision Dilemma: A Case of Entanglements and Irreconcilable Differences." *Journal of Curriculum and Supervision* 9 (2): 195–216.

Heller, Marjorie F., and William A. Firestone. 1995. "Who's in Charge Here? Sources of Leadership for Change in Eight Schools." *Elementary School Journal* 96 (1): 65–86.

Hewitt, Denise. 2000. "The Cincinnati Plan." In *The Peer Assistance and Review Reader*, ed. Gary Bloom and Jennifer Goldstein. Santa Cruz, CA: New Teacher Center.

Hunter, Madeline. 1988. "Effecting a Reconciliation between Supervision and Evaluation—a Reply to Popham." *Journal of Personnel Evaluation in Education* 1 (3): 275–79.

Iwanicki, Edward F. 1998. "Evaluation in Supervision." In *Handbook of Research on School Supervision*, ed. Gerald R. Firth and Edward F. Pajak. New York: Macmillan.

Kaboolian, Linda, and Paul Sutherland. 2005. "Evaluation of Toledo Public School District Peer Assistance and Review Plan." Unpublished report, John F. Kennedy School of Government, Harvard University, Cambridge, MA.

Kelly, Philip P. 1998. "Teacher Unionism and Professionalism: An Institutional Analysis

of Peer Review Programs and the Competing Criteria for Legitimacy." PhD diss., Michigan State University.

Kerchner, Charles Taylor, and Julia E. Koppich. 1993. *A Union of Professionals: Labor Relations and Education Reform*. New York: Teachers College Press.

Kerchner, Charles Taylor, Julia E. Koppich, and Joseph G. Weeres. 1997. *United Mind Workers: Unions and Teaching in the Knowledge Society*. San Francisco: Jossey-Bass.

Kerchner, Charles Taylor, and Douglas E. Mitchell. 1988. *The Changing Idea of a Teachers' Union*. New York: Falmer.

Koppich, Julia E., and Charles Taylor Kerchner. 1999. "Developing Careers, Building a Profession: The Rochester Career in Teaching Plan." Paper prepared for the National Commission on Teaching and America's Future.

Lankford, Hamilton, Susanna Loeb, and James Wyckoff. 2002. "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis." *Educational Evaluation and Policy Analysis* 24 (1): 37–62.

Leithwood, Kenneth, and Doris Jantzi. 2000. "The Effect of Different Sources of Leadership on Student Engagement in School." In *Leadership for Change and School Reform: International Perspectives*, ed. Kathryn A. Riley and Karen Seashore Louis. New York: Routledge/Falmer.

Little, Judith Warren. 1988. "Assessing the Prospects for Teacher Leadership." In *Building a Professional Culture in Schools,* ed. Ann Lieberman. New York: Teachers College Press.

Lortie, Dan C. 1975. *Schoolteacher*. Chicago: University of Chicago Press.

Loup, Karen S., Joanne S. Garland, Chad D. Ellett, and John K. Rugutt. 1996. "Ten Years Later: Findings from a Replication of a Study of Teacher Evaluation Practices in Our 100 Largest Districts." *Journal of Personnel Evaluation in Education* 10 (3): 203–26.

McGreal, Thomas L. 1997. "Can a Supervisor Be a Coach? Yes." In *Educational Supervision: Perspectives, Issues, and Controversies*, ed. Jeffrey Glanz and Richard F. Neville. Norwood, MA: Christopher-Gordon.

Miles, Matthew B., and A. Michael Huberman. 1994. *Qualitative Data Analysis*. Thousand Oaks, CA: Sage.

Murray, Christine E. 1999. "Rochester Teachers Struggle to Take Charge of Their Practice." In *Transforming Teacher Unions*, ed. Bob Peterson and Michael Charney. Milwaukee: Rethinking Schools.

NCTAF (National Commission on Teaching and America's Future). 1996. *What Matters Most: Teaching for America's Future*. New York: NCTAF.

Nolan, James F. 1997. "Can a Supervisor Be a Coach? No." In *Educational Supervision: Perspectives, Issues, and Controversies*, ed. Jeffrey Glanz and Richard F. Neville. Norwood, MA: Christopher-Gordon.

Ogawa, Rodney T., and Steven T. Bossert. 1995. "Leadership as an Organizational Quality." *Educational Administration Quarterly* 31 (2): 224–43.

Painter, Suzanne R. 2000. "Principals' Perceptions of Barriers to Teacher Dismissal." *Journal of Personnel Evaluation in Education* 14 (3): 253–64.

Peterson, Kent D. 1995. *Teacher Evaluation: A Comprehensive Guide to New Directions and Practices*. Thousand Oaks, CA: Corwin Press.

Popham, W. James. 1988. "The Dysfunctional Marriage of Formative and Summative Teacher Evaluation." *Journal of Personnel Evaluation in Education* 1 (3): 269–73.

Smith, Thomas M., and Richard M. Ingersoll. 2004. "What Are the Effects of Induction and Mentoring on Beginning Teacher Turnover?" *American Educational Research Journal* 41 (3): 681–714.

Smylie, Mark, Sharon Conley, and Helen M. Marks. 2002. "Exploring New Approaches to Teacher Leadership for School Improvement." In *The Educational Lead-*

ership Challenge: Redefining Leadership for the 21st Century, 101st Yearbook of the National Society for the Study of Education, pt. 1, ed. Joseph Murphy. Chicago: NSSE.

Spillane, James. 2006. *Distributed Leadership*. San Francisco: Jossey-Bass.

Spillane, James P., Richard Halverson, and John B. Diamond. 2001. "Investigating School Leadership Practice: A Distributed Perspective." *Educational Researcher* 30 (3): 23–28.

Strong, Michael. 2004. "Induction, Mentoring, and Teacher Retention: A Summary of the Research." Santa Cruz, CA: New Teacher Center.

Stronge, James H. 1995. "Balancing Individual and Institutional Goals in Educational Personnel Evaluation: A Conceptual Framework." *Studies in Educational Evaluation* 21: 131–51.

Sykes, Gary. 1987. "Reckoning with the Spectre." *Educational Researcher* 16 (6): 19–21.

Tucker, Pamela D. 1997. "Lake Wobegon: Where All Teachers Are Competent (Or, Have We Come to Terms with the Problem of Incompetent Teachers?)." *Journal of Personnel Evaluation in Education* 11 (2): 103–26.

Tyack, David B. 1974. *The One Best System*. Cambridge, MA: Harvard University Press.

Urbanski, Adam. 1999. Presentation at the annual meeting of the Teacher Union Reform Network, Santa Cruz, CA, November.

Van Maanen, John, and Stephen R. Barley. 1984. "Occupational Communities: Culture and Control in Organizations." *Research in Organizational Behavior* 6:287–365.

Wasley, Patricia A. 1991. "The Practical Work of Teacher Leaders: Assumptions, Attitudes, and Acrophobia." In *Staff Development for Education in the '90s: New Demands, New Realities, New Perspectives*, ed. A. Lieberman and L. Miller. New York: Teachers College Press.

Wise, Arthur E., Linda Darling-Hammond, Milbrey W. McLaughlin, and Helen T. Bernstein. 1984. *Teacher Evaluation: A Study of Effective Practices*. Santa Monica, CA: RAND.

Yin, Robert K. 2003. *Case Study Research: Design and Methods*. 3rd ed. Thousand Oaks, CA: Sage.