JPART 20:23–52

Re-Creating Street-Level Practice: The Role of Routines, Work Groups, and Team Learning

Erica Gabrielle Foldy
New York University
Tamara R. Buckley
City University of New York

ABSTRACT

Ample research documents the ubiquity of routines in street-level practice. Some individual-level and organizational-level research has explored how to break street-level routines, but little has looked at the work group level. Our study observed teams of state child welfare workers over 2.5 years, documenting whether they discarded old routines and learned new ones. Results suggest that team characteristics such as clear direction and reflective behaviors had greater influence on team learning than individual characteristics such as stress level, tenure, and educational level. We suggest that group-level factors be included in future models of what enables the re-creation of street-level practice.

Ample research suggests that street-level practice in public organizations tends to be routinized, inflexible, and difficult to change (Lipsky 1980; Maynard-Moody and Musheno 2003; Meyers, Glaser, and Mac Donald 1998; Riccucci 2005). Still, we do have some data on the factors that enable changes in practice, including individual-level variables like education and experience and organizational-level factors like incentive and reward systems (Jewell and Glaser 2006; Riccucci et al. 2004). However, little of this work has explored the *workgroup* level and its role in change and learning in street-level practice. This is curious because extensive literature in social psychology and organizational behavior has investigated how groups break old routines and learn new ones (e.g., Edmondson 1999; Gersick and Hackman 1990).

This article draws on research in a child welfare agency that has been experimenting with using teams of social workers, rather than individual workers, to engage with families involved in the child protection system. We found that team characteristics, such as clear sense of direction and capacity for reflection, seem to enhance their members' ability to

The authors thank the editor and anonymous reviewers for their very helpful comments as well as the members of Fulton 214 (Pacey Foster, Danna Greenberg, Tammy MacLean, Peter Rivard, Jenny Rudolph, and Steve Taylor). Very special appreciation goes to the Massachusetts Department of Children and Families and the employees, at every level, who made this research possible, and to the Russell Sage Foundation for its support. Address correspondence to the author at erica.foldy@nyu.edu.

doi:10.1093/jopart/mun034
Advance Access publication on February 10, 2009

learn new routines, suggesting that research at the team level could be important for understanding how to change street-level practice.

We begin by reviewing, in more depth, the literature on frontline practice, workgroups and team learning, and then present hypotheses. Following that, we summarize the methods used in this research and present our findings. Finally, we discuss our results and their contribution to the literature on change and learning in street-level practice.

STREET-LEVEL PRACTICE: ROUTINES, WORKGROUPS, AND LEARNING

A significant literature has documented the presence of embedded routines, and the difficulty of changing those routines, in public organizations. For example, the extensive research on whether and how welfare reform affected agency practice generally found more stasis than change across multiple states (Jewell and Glaser 2006; Meyers, Glaser, and Mac Donald 1998; Meyers, Riccucci, and Lurie 2001; Riccucci 2005; Riccucci et al. 2004; Sandfort 1999). However, this and other research has identified some individual and organizational elements that can make a difference. Riccucci, Meyers, and their colleagues found that agency practices including clear goals, training, and monitoring did encourage employees to change their perceptions of the purpose of welfare reform (Meyers, Riccucci, and Lurie 2001; Riccucci et al. 2004). Jewell and Glaser (2006), also investigating the arena of welfare reform, found a number of elements that inhibited change, such as workload, role expectations, incentives, and worker expertise. They argue that significant restructuring by having workers specialize in different tasks was the most effective course. Yin (1981) identified the support of top-level administrators as crucial in technological innovation. However, this work on modifying street-level practice has largely overlooked factors at the workgroup level (an exception is West and Poulton 1997).

A related body of work focusing on street-level workers more generally does reference the importance of coworkers in determining how work is carried out. Vinzant and Crothers (1998, 11) name coworkers as 1 of 10 "influences on street-level public servants." Sandfort (1999) points to the importance of "collective schema" held by frontline workers which can inhibit change and learning. However, coworkers can also be a source of learning. Maynard-Moody and Musheno (2003, 160) argue that self-reflection is critical to breaking out of embedded assumptions and ways of working and that connecting with peers is more likely to create the conditions for self-reflection. Lipsky (1980, 209) also cites the importance of peer discussion, particularly "ongoing processes of supportive criticism and inquiry" that truly challenge individual practice. However, none of this work actually investigated groups of workers, as opposed to individual workers, and, thus, did not systematically explore group characteristics and how they affect how work is done.

Yet, significant social psychology and organizational behavior research attests to the importance of work group processes and characteristics in changing practice. Much of this work is about group *effectiveness* (Campion, Medsker, and Higgs 1993; Goodman, Ravlin, and Schminke 1987; Guzzo and Dickson 1996; West and Poulton 1997), but more recently a related literature on team *learning* has appeared (Argote, Gruenfeld, and Naquin 2001; Edmondson 1999; Gibson and Vermeulen 2003; Scarbrough, Bresnen, et al. 2004, Scarbrough, Swan, et al. 2004; Wenger 2000). This shift in focus acknowledges that organizations and groups must become adaptable and flexible, given a rapidly changing environment.

Team Learning

Learning has been defined both as a process and as an outcome (Edmondson and Moingeon 1998; Edmondson, Dillon, and Roloff 2007). Those defining it as a process construe it as particular behaviors, such as raising errors or asking for help (e.g., Gibson and Vermeulen 2003; Tamuz 2001). Those defining it as an outcome have defined it as a change in knowledge (e.g., Huber 1991; Zajac and Comfort 1997), in actions and routines (Ellis et al. 2003; Freeman 2007), or in both (Dekker and Hansen 2004). Given the interest of public management scholars and practitioners in how to "unfreeze" (Lewin 1947) embedded routines in street-level practice, defining learning as a change in routines is the most relevant and robust approach. Therefore, we use the influential definition of Levitt and March (1988, 320) because it suggests that learning occurs when organizations establish new routines: "[O]rganizations are seen as learning by encoding inferences from history into routines that guide behavior." They define routines as "beliefs, frameworks, paradigms" as well as "forms, rules, procedures" (1988, 320). This definition also has the benefit of encompassing both knowledge and behavior, which is particularly important when studying professionals (Currie, Finn, and Martin 2007; Currie and Suhomlinova 2006).[1]

Public management researchers have begun exploring learning in the public sector, with significant work at the organizational level (e.g., Dekker and Hansen 2004; Mahler 1997) and some at the interorganizational (Rashman and Hartley 2002) and individual (Goldman 2007; Hill 2003) levels. There is little public management research on team learning per se. However, there is work on closely related topics such as communities of practice (Wenger 2000; Wenger and Snyder 2000) and project-based learning (Scarbrough, Bresnen, et al. 2004, Scarbrough, Swan, et al. 2004) which, along with other management research on team learning and effectiveness, provides useful guidance on what factors might enable teams of street-level workers to replace embedded practices with new ways of working.

Prior Research on Team Characteristics

Prior research suggests a number of team characteristics critical for team learning and effectiveness. This research has taken place largely in business organizations (exceptions include the work of West and colleagues [e.g., West 1996; West and Poulton 1997]), which provide a different environment from governmental organizations. For example, some research suggests that public employees are more risk averse, and have lower job satisfaction and organizational commitment (Boyne 2002; Buchanan 1974), and that public organizations are less likely to reward employees for enhancing their own skills and for innovation (Boyne, Jenkins, and Poole 1999). All these qualities could influence teamwork and team outcomes. Furthermore, there is much less research on team learning, even in business organizations, so it is less clear whether these critical characteristics play a similar role. Given these uncertainties, we had a unique opportunity to see whether previous findings would hold up in a study of team learning in public organizations.

Drawing on previous research on both team effectiveness and learning, we developed hypotheses on the team characteristics that could have an impact on team learning.

1 A second debate in the learning literature centers on whether learning is defined descriptively or normatively. Normative definitions presume that learning is always positive, whereas descriptive approaches suggest that learning can be both good and bad (Vera and Crossan 2005). Given the interest in what loosens routines in street-level practice, in this article, we presume that learning that unfreezes standard practice is positive, recognizing that that may not be true in all circumstances.

However, given that prior research suggests some individual-level variables can also be influential, we investigated those as well.[2]

Team-Level Characteristics

Research suggests a number of important characteristics. Clarity on two different dimensions appears to be a prerequisite. The first dimension is the group's goals: a clear and shared direction—a sense of purpose or mission—has been shown to influence team effectiveness: "[T]here does appear to be a strong evidentiary basis for the performance effects of goals" (Guzzo and Dickson 1996, 314). Goals provide both the motivation for joint work as well as a focus for teams to refer to as they negotiate how to share and coordinate their work (Guzzo and Dickson 1996; West and Poulton 1997).

The second dimension is clarity regarding the team's composition or a clear sense of membership. The importance of boundaries was identified early in the organizational behavior literature: "The small face-to-face group provides a boundary within which the member can be known and can feel secure; within which . . . he can seek reinforcement and help" (Miller and Rice 1967, 19) and continues to be relevant (Scarbrough, Bresnen, et al. 2004, Scarbrough, Swan, et al. 2004). Teams with a clear sense of their membership not only understand who should be involved in decision making and in sharing the work but feel a greater sense of safety (Gladstein 1984; Wageman, Hackman, and Lehman 2005).

This summary suggests the following two hypotheses:

H_1 Clear direction will enhance team learning.

H_2 Clarity of team composition will enhance team learning.

A group's sense of stability and cohesion is also important for both effectiveness and learning. Two elements are important here. First, team turnover (or the lack of stability) has been shown to interfere with group effectiveness and learning (Argote, Gruenfeld, and Naquin 2001; Edmondson, Bohmer, and Pisano 2001; West 1996).[3] Team stability is considered crucial because it can enhance the creation of shared mental models that facilitate communication and knowledge retrieval (Edmondson, Dillon, and Roloff 2007, 282). Team turnover could be related to a fourth factor, a group's social support (Campion, Medsker, and Higgs 1993) or cohesion (Guzzo and Dickson 1996). Mutual support is important for developing the trust required for members to rely on each other and work collectively. The literature on communities of practice emphasizes this dimension of teamwork: A community's effectiveness depends, in part, on mutuality: "[Community members] must trust each other, not just personally, but also in their ability to contribute to the enterprise of the community, so they feel comfortable addressing real problems together and speaking truthfully" (Wenger 2000, 230).

This summary results in these hypotheses:

H_3 Team turnover will diminish team learning.

H_4 Mutual support will enhance team learning.

2 Some work also proposes organizational-level factors as important. These were beyond the scope of our study since our teams were all part of one organization, but future research should explore this arena.

3 Team turnover may seem similar to clarity of team membership, but it is not the same thing. Clarity of composition refers to the roles or job positions that are considered part of the team's make up. Team turnover refers to whether the occupants of those roles or positions remain the same or change over time.

Another area of interest is a team's leadership, both internal and external. Internally, the team leader plays a central role in the group (Edmondson 2003; Goodman, Ravlin, and Schminke 1987; Guzzo and Dickson 1996). The important role of supervisors in the public sector has already been established (Brewer 2005; Riccucci et al. 2004). However, they play a special role in teamwork since their capacity to help set direction, coach members, coordinate with other employees outside the team, and facilitate discussion and debate could all affect a team's capacity to learn (Wageman, Hackman, and Lehman 2005). The role of a team's superiors can also be crucial since they can support or undermine a team's capacity to work together (Edmondson, Bohmer, and Pisano 2001; Hackman 1987). They do this by helping to create the team's purpose and relating it to the organization's mission, by creating appropriate reward systems, and by connecting the team's work to the rest of the organization, including running interference as necessary (Miller and Rice 1967, 20–21; Snyder et al. 2003; Wenger and Snyder 2000). These hypotheses follow from this discussion:

H_5 Team leader effectiveness will enhance team learning.

H_6 Support from the team's superiors will enhance team learning.

Two team behaviors have been linked to *learning* in particular (Argyris 1990; Edmondson 1999). First is the capacity of the team to experiment with new ways of doing the work. This would seem straightforward: If a team sticks to its traditional routines, it is impossible for it to learn new ones. Second, a team's propensity to reflect on its work, by surfacing and discussing embedded routines, by raising errors, and by engaging differences of opinion should also be critical (Edmondson 1999; Edmondson et al. 2001). Work on project-based learning suggests that reflection is fundamental to learning (Scarbrough, Bresnen, et al. 2004), and research on communities of practice has identified self-awareness as crucial in helping such groups be "cradles of the human spirit" rather than its "cages" (Wenger 2000, 230).

This research has the following implications:

H_7 Experimentation will enhance team learning.

H_8 Reflective behaviors will enhance team learning.

Individual-Level Characteristics

Although our focus is on team-level characteristics, we investigated whether individual-level factors might be important as well, in order to test alternate explanations for our results as required for case study research (Yin 2003). Teams are made up of small numbers of individuals and of course will be affected by the characteristics of their members (see, e.g., West 1996). Previous research suggests several important individual-level variables. First, stress has been shown to diminish flexibility and strengthen rigidity (Staw, Sandelands, and Dutton 1981), which is likely to undermine the capacity to create new routines. Frontline public service work is often stressful and child welfare work is at the high end of the continuum (Nissly, Mor Barak, and Levin 2005). Therefore, it was critical to see if higher stress levels hampered learning.

Second, worker skills and experience can also influence team effectiveness (Argote, Gruenfeld, and Naquin 2001; Campion, Medsker, and Higgs 1993). The literature suggests two important aspects of experience and expertise: job tenure and educational level

(Gladstein 1984; Riccucci et al. 2004). Research on street-level bureaucrats suggests that length of job tenure could work against the breaking of routines since employees with more years on the job could be more entrenched in the traditional ways of working and resistant to change (Lipsky 1980, 143–144).[4] Regarding education level, the research thus far suggests that more education is likely to increase learning (Gladstein 1984; Vera and Crossan 2005). This is especially true given that the complex nature of child welfare work would seem to require strong clinical and casework skills as well as keeping abreast of developments in the field.

Finally, job commitment and job satisfaction have been shown to affect many aspects of work since they are linked to work motivation (George and Jones 1997; Millward and Hopkins 1998; Moynihan and Pandey 2007). Learning new ways of working requires greater interest and energy than sticking with old routines. This summary suggests the following hypotheses about individual-level characteristics:

H_9 Higher stress will diminish team learning.

H_{10} Shorter tenure will enhance team learning.

H_{11} Higher educational level will enhance team learning.

H_{12} Higher job satisfaction and job commitment will enhance team learning.

METHODS

Our research followed seven teams of social workers over several years, documenting how they modified their work and whether those changes were maintained. These workers were employees of a state child protection agency experimenting with a team-based structure because of concern about the stressful working conditions facing its frontline employees. The study used a comparative case study design, and both qualitative and quantitative methods, to answer the question: What enables street-level workers to break old routines and learn new ones?[5]

Research Site

In 2004, the Massachusetts Department for Social Services (as of July 2008, the agency's name was changed to Department of Children and Families), a state child protection

4 Tenure is related to team turnover, but tenure is an individual-level variable, whereas team turnover is team level. One could potentially have a team comprised of people who have many years of experience, but is unstable due to members leaving and new members coming in. Even if those new members also have many years of experience, that would not necessarily lessen the effect of the turnover.

5 We were not evaluating the outcomes of teams of social workers compared with individual social workers. Rather, we focused on the seven teams to see if we could determine if and how the teams were able to break with long-established child welfare practice which relies on a single worker and her therapeutic relationship with a family. Although the agency ultimately declared the experiment a success and is in the process of disseminating team practice more broadly, we are agnostic about whether teaming improved family outcomes since that was not the focus of our research. However, given the keen interest of both researchers and practitioners in how to dislodge bureaucratic routines and learn new ways of working, further insight into that question is important in itself. This is particularly true regarding the adoption of team-based structures since they have been a popular innovation within the public sector, both in the United States and abroad (Leisink 2004; Hickman and Creighton-Zollar 2001).

agency, launched a pilot Teaming Initiative to explore how a team-based structure could improve social worker morale as well as services to families. The agency was concerned about its frontline employees who work directly with clients and suffer considerable stress and isolation (Nissly, Mor Barak, and Levin 2005). It hoped that a team-based model could facilitate better outcomes for workers and clients in two ways: by bringing together a broader set of perspectives and experiences in thinking about how to work with a given family as well as by providing more emotional support to individual social workers.

These frontline workers were the agency employees out in the field, going to people's homes to investigate allegations of child abuse or neglect, speaking with teachers, counselors, and other "collaterals" as the agency called them, finding resources (such as counseling or temporary financial support) for families, and making recommendations to their superiors about whether or not a child should be removed from a home.

The agency had divided the sequence of working with a family into three "functions": *investigation*, or the task of following up on allegations of abuse or neglect and determining whether to take on the case; *assessment*, or the job of doing a more extensive exploration of the family and creating a "service plan" for the family; and *ongoing*, the work of implementing the service plan and engaging with the family to ensure a safer environment for the child.

In the traditional set-up, each social worker was a member of a "unit" with a supervisor and four or five workers. For the most part, a unit engaged in only one of the three functions and, thus, was known as either an investigation unit, an assessment unit, or an ongoing unit. Members of these units sat near each other, but each worker had his or her own caseload and was singly responsible for those cases. (Caseload was officially capped at 18, but often went higher.)

In the Teaming Initiative, the structure was similar: each worker was a member of a "team" with a supervisor and four or five social workers. However, the form of service delivery was expected to change in that multiple workers would be involved in cases and all team members could be called on for advice and assistance. Beyond these general parameters, the agency left it up to individual teams to define in more depth what cases they would focus on and how they would structure their work. Participation in the Initiative was voluntary: individual office directors decided if they wanted one of their units to participate in the Initiative and, if so, asked their units who might be willing. Interested offices sent in a proposal in which they described how their team members would work together and for what purpose. This resulted in wide variation in how the teams implemented the concept of "teaming," which provides an excellent opportunity to investigate how their differences may have affected their capacity to learn.

Data Collection

The Teaming Initiative involved seven teams in six offices around the state.[6] The first author enrolled all seven teams in the research and collected both qualitative and quantitative data, working with a research assistant, a doctoral student in social work.

6 There were eight teams initially but one disbanded within a couple of months because the leader chose to move to a different office.

We used a comparative case study design (Agranoff and Radin 1991; Eisenhardt 1989; Yin 2003) for several reasons. First, case studies are appropriate when the interest is in "operational links needing to be traced over time" (Yin 2003, 6). We were curious if, over time, some teams would learn better than others and, if so, what enabled them to do so. Second, the Teaming Initiative at the agency involved only a small number of teams, yet, given findings from prior research, a number of potential factors could be at play. Comparative case study designs are useful for situations in which "there will be many more variables of interest than data points" (Yin 2003, 13). Case study research is often used for inductive rather than deductive research (Agranoff and Radin 1991) but can be used to hone existing theory (Siggelkow 2007). Given the small sample, however, we have to be cautious about any conclusions we draw.

Case study research requires explicit testing of alternative explanations for the observed phenomenon in order to enhance the robustness of the findings. Although our focus is on team-level characteristics and their relationship to team learning, we also investigated individual-level variables to see whether these might also have a causal relationship with team learning, given previous team research.[7]

Data collection began in summer, 2004, and finished in the winter of 2007. Three rounds of data collection took place: at the beginning of the teams' work together (2004), after about 1 year (2005), and after 2–2.5 years (2006−07). We gathered both qualitative and quantitative data during each round, including surveys with team members and team interviews (interviewing the whole team at one time). We also conducted surveys and interviews with team stakeholders and observers, including the team's immediate superior, the director of the office in which the team was located, a teaming consultant working with the teams, and the coordinator of the Teaming Initiative. These two latter stakeholders were familiar with all teams. This article draws from the team surveys and interviews from Rounds 1, 2, and 3 and interviews with team stakeholders from Rounds 2 and 3 for a total of 28 interviews and 114 surveys.

The team interviews gathered data on how the teams were structuring their work together and how this differed (or did not) from traditional casework. Because we interviewed the teams three times over about 2.5 years, we could see how their practices changed (or did not) over time, and we could draw conclusions about whether the teams had truly engaged in team learning, in that they had changed from working autonomously to working collectively. The interviews with the stakeholders solicited their opinions on the successes and challenges of the teams they were familiar with, allowed us to confirm or disconfirm claims by the teams, and provided additional data and perspectives on our key constructs. Finally, the survey data included information on individual-level data, including stress, job satisfaction, job commitment, tenure at the organization, and educational level.

We gathered data on 14 different constructs. Our outcome variable, or the phenomenon we are trying to explain, is *team learning*, defined as the extent to which the teams discarded old routines and embedded new routines in their work together. As Lewin (1947, 34) notes, change comes in three steps—unfreezing, moving, and refreezing—and if the

7 The type of replication is also important in comparative case study design (Yin 2003): do the cases provide literal replication (that comes from similarity across cases) or theoretical replication (that arises from differences across cases)? Our full sample represents both literal and theoretical replication, in that the seven teams ultimately fell into three categories, each with two or more teams, based on their capacity for learning. This created the variance in category required for theoretical replication and the similarity within category required for literal replication.

change does not refreeze, change has not occurred. Therefore, we paid careful attention to whether attempted changes in fact became rooted ones (over about 2.5 years).

The agency wanted the workers to stop working individually on cases and begin working collectively. Therefore, we looked for three specific changes in routines to establish whether team learning had occurred: (1) Did the team meet regularly (at least weekly) as a full group to review cases and discuss team issues? (2) Did team members, on an ad hoc basis, ask for and provide help on each others' cases? (3) Did the team assign at least two workers to take formal, joint responsibility for a case?[8] In this way, our operationalization of "team learning" is, in a sense, "learning to team": for learning to have occurred, it had to happen on the team level, not the individual level. (An individual member can not team by herself.)

We assessed the extent of team learning from team interview questions about how team members worked on cases, whether they asked for and received help, whether and how they shared work, and how often they met and the kinds of things they discussed. We then determined whether the team members had made changes in their work practice (compared with the traditional way of working) and whether these changes had been maintained for the first 2.5 years the team worked together. We used the team stakeholders to confirm their assertions. The key element to determining team learning was *whether changes were maintained*, not the nature of the particular changes. We had 13 criterion, or explanatory, variables; here we define each one and explain how it was assessed:

Team Direction (H_1)

The extent to which the team felt it had a clear, shared objective. Assessed using team interview questions including why they had decided to become a teaming unit, the kinds of casework they engaged in, and how they defined a "teaming" case. Used the stakeholder interviews to gather additional information.

Clarity of Team Composition (H_2)

The degree to which team members were clear about the team's boundaries or who was and was not a team member. Assessed with a team interview question about whether team membership was clear.

Turnover (H_3)

The extent to which members entered and left the team over the first 2.5 years together. Assessed each round at the team interview. (There are a variety of turnover measures; we calculated the "separation rate," or the rate at which team members left the team, a common measure [Price 1976].)[9]

Mutual Support (H_4)

The degree to which team members trusted and felt they could count on each other. Assessed through team interview questions about how they divided up the work, how they

8 We also assessed whether other new routines had been established, for example, in the functions the teams engaged in or the timing of services given to families, but not all teams attempted these changes, so we could not systematically compare these kinds of changes across teams.

9 Turnover can also be measured using the accession rate, the stability rate, and the instability rate (Price 1976). We calculated all these measures, but they did not change the teams' ranking using the separation rate.

communicated with each other, what they talked about in meetings, and the benefits and challenges of teaming. Used the stakeholder interviews for additional data and perspectives.

Team Leader Effectiveness (H_5)
The degree to which team members and stakeholders felt the team leader (whose job title was ''supervisor'') enabled the team to work together successfully. Assessed through stakeholder questions about the team leader's management style and whether the team leader was responsible for the team's level of effectiveness. Also, although we could not ask about the team leader directly in the team interviews because she or he was always present, we did get some data that suggested how the leader was seen by the members.

Support from Superiors (H_6)
The extent to which team members felt their superiors enabled them to carry out their work. Assessed using team interview questions about the larger organizational context and whether they felt they were getting the resources and support to do their work. Used the stakeholder interviews to gather additional data and perspectives.

Experimentation (H_7)
The extent to which team members experimented with new ways to implement their work. Assessed using team interview questions about how the team was carrying out its work and whether they had experimented with different approaches. Used stakeholder interviews to confirm as well as gather additional information.

Reflection (H_8)
The extent to which team members considered how they were carrying out the work and how it could be done differently or better. Assessed using team interview questions about whether they had set aside time from actual task work to consider new or better ways of proceeding with the work and whether they ever brought up difficult issues, asked for feedback, or raised concerns about mistakes or errors. Used stakeholder interviews to confirm as well as gather additional information.

Stress (H_9)
The degree to which team members experienced both time stress and anxiety stress. Assessed with survey, using a12-item measure from Parker and Decotiis called the Job Stress Scale (Fields 2002, 132), with a scale where 1 was low and 5 was high.

Worker Tenure (H_{10})
The length of job tenure at the agency. Assessed with survey question about how long they had been working at the agency. We converted their answers to months.

Worker Education (H_{11})
The highest level of education completed. Assessed with survey question. Since all workers had either a bachelors or a masters, we developed a measure of the percentage of team members with a master's degree.

Job Commitment (H_{12})
The extent to which members felt psychological commitment to their job and the agency. Assessed with survey, using 10-item measure developed by Tsui, Egan, and O'Reilly (1992), with a scale where 1 was low and 5 was high.

Job Satisfaction (H_{12})

The extent to which members felt satisfaction with different aspects of their job, including their pay, supervisor, etc. Assessed with six-item Job Satisfaction Index developed by Schriesheim and Tsui (Fields 2002, 162), a scale where 1 was low and 5 was high.

Data Analysis

We came to this analysis starting from very different positions. The first author had gathered most of the data; the second author had no previous involvement with the study and therefore came with fresh eyes. We used these different positions as a way to deepen our analysis and check our assumptions.

Our overall analysis strategy was both descriptive and analytic. We developed a narrative or case description for each team, which tracked the team's progress over its first 2.5 years together. We also did an analysis across all seven teams, looking for evidence in each team of the specific characteristics we cite above. The type of analysis varied, depending on whether we were working with qualitative or quantitative data.

Qualitative Data

We began the process by both reading and coding all the transcripts for the Round 2 team interviews (a total of seven, one for each team). As a process check, we independently categorized the teams as low, medium, or high on the outcome variable, their learning, and on several of the key explanatory characteristics. For example, we looked for evidence of whether the team had a clear and shared direction and then categorized it as either low, medium, or high in Direction. We achieved an initial agreement rate of 74%. (Please see table 1 for illustrations of the evidence used to categorize teams as low, medium, or high on constructs measured with qualitative data.)

We then both read the Round 1 and Round 3 interviews for four of the seven teams and then independently categorized those four teams as low, medium, or high on learning and on the criterion constructs. If the team seemed to change over time on a given construct, we noted that; otherwise we gave them an overall categorization. For those that did change over time, we looked to see whether that change seemed to have a significant impact on our outcome variable of team learning; if not, we created an overall categorization that took into account the team's variance over time.

In this round, we achieved an agreement rate of 80%, which is quite acceptable (Boyatzis 1998, 156). When we differed, we each identified the particular data points which had influenced our assessment, clarified whether they were actually germane to the construct, and then discussed how to interpret that data. The results reported here reflect our joint understanding. Based on that rate of agreement, only the first author read the material and categorized the other three teams.

We then used pattern matching, an analytic method that compares an "empirically based pattern with a predicted one" (Yin 2003, 116). In our case, this meant comparing the categorization on a given criterion or explanatory variable with the categorization on the outcome variable, which was team learning. A perfect match would be that the categorization of the teams as low, medium, or high on the criterion variable exactly matched their categorization as low, medium, or high on learning, but perfect matches are not always necessary (Eisenhardt and Graebner 2007, 29). We used agreement in six or seven of the seven teams as the basis for drawing preliminary conclusions.

Table 1
Illustrations of Evidence Used to Categorize Teams on Constructs Measured with Qualitative Data

	Low	Medium	High
Direction (H_1)	R: The reason we stopped [teaming] is because there was no involvement and direction as to how it . . . needed to look in this office . . ..With [the director of their office], there was no direction as to "what do you want this look like?"	R: I think we need some time to reflect a little where we're headed because we're kind of regrouping with new members. We got to take out our mission statement again—and just reflect and figure out where we're headed from this point on.	A member of one team described how, through teaming, its members can interact with the family in a "strength-based," more supportive approach, which was their goal in undertaking a team approach. "[Teaming] is a much better approach because we have the time now to work on some strength-based stuff that we never used to. We used to just go out, do investigations. . . and it was just strictly adhere to. . . what's going wrong with the family. . . And now we have the option of saying 'What is going good? Who do you have for support? What can we do for you?' It's much more strength-based and much more relaxed than the typical going out and investigating with the authoritative type style. We don't work like that anymore."

Continued

Table 1 *(continued)*
Illustrations of Evidence Used to Categorize Teams on Constructs Measured with Qualitative Data

	Low	Medium	High
Clear composition (H_2)	INT: So, would you say, at this point, that the membership of the team is clear? And if so, who's a member of the team? [long pause] R: I think it's clear. INT: So, who's on the team? R: Everybody in this room, including myself. INT: And, anybody who's not in the room? . . . R: Well, the Area Director. R: She hasn't attended our meetings for a very long time. R: She's only been at a couple. . . . R: Well, if she's not here, it doesn't – to me, it doesn't feel like she's part of the team. . . . R: But at the same time, she is a member of it because . . . she does have a voice in . . . saying what the final say is. So, in that way, I do consider her to be a member of the team.	INT: Are you totally clear who is a member on this team? R: Yes. INT: Who are the members? R: [gives names]. INT: Does anybody else ever come in and sit in on meetings? . . . R: Yes, other people have been involved in the teaming—There was a brief time when there was actually an extra person assigned to Peter's unit that joined the team for a couple of months . . . INT: So, people may come and take part in the meeting? R: Yes.	INT: Is the team membership clear? Do you know who is on the team? R: Yes. R: Definitely. INT: Okay and is that everybody in this room? R: Yes. INT: Anybody who is not in the room? R: No. INT: So sometimes teams have like a core membership and then a larger membership and it sounds like you guys don't— R: Not at this point.

Continued

Table 1 *(continued)*
Illustrations of Evidence Used to Categorize Teams on Constructs Measured with Qualitative Data

	Low	Medium	High
Mutual support (H_4)	R: That becomes the greatest problem with teaming. It's much easier not to say anything. I know what I want to do. I'm not even going to bring anything to the team. . . . But I think . . . everybody does that, not just me. It's easier . . . INT: Would you all say you might bring more things to the team than [name of respondent who just spoke] does? R: I have no idea. We haven't really worked it out. . . . R: I'm going to pick a random week. I'm going to keep track of everybody and how often do they take something to the entire team? I guarantee it's not more than one time per person, once a week. R: I wouldn't necessarily agree with that but I don't take anything to the team, actually . . . R: I don't know how much I take to the team, either.	R: (to another member of the team, who has just said that, during a recent emergency, she had no team members to talk to) Would you feel uncomfortable talking to [she names two other team members] and I? R: No. R: Because we were there [when you needed help]. I don't— R: I just think that, I definitely do have people that I'm more comfortable with and I just feel like you guys are kind of like newer . . . R: That's interesting to look at, because we are all part of the team.	R: I keep going by the unit and I see [name of team member] sitting with [another name of member] at her desk, and they are literally sitting this close to each other looking at the computer doing all this work together. . . . Or [name of member] . . . went to a home alone and handled a situation—that was a terrible situation—beautifully, had a conversation with [name of member] about it and then they were in my office saying, "This is a safety issue and we need to you know."

Continued

Table 1 *(continued)*
Illustrations of Evidence Used to Categorize Teams on Constructs Measured with Qualitative Data

	Low	Medium	High
Team leader effectiveness (H_5)	One office director said that the team leader in her office was hired "out of the agency" and was "never fully integrated. And her own style, personality is such—she hasn't made an effort to integrate herself into that group [her team]. [She] created stress and work [that] they [the team] didn't have time to do." Later she noted that the team leader was "lacking." When asked why, the director described the leader's "style" as "tenacious, rigid, controlling."[a]	One office director described the team leader in his office as "wonderful" with a "good organizational capacity" but also noted she was engaged in a, "'am I the co-worker or am I the supervisor?' kind of struggle and I think she's really struggling with that piece . . ." Later, he noted, "[The team leader] takes on [a particular] case, doesn't really have any role . . . People are perceiving her as meddling in other people's work."	One office director noted that the team leader in his office "really pushes workers to do the best they can, for themselves and for the families," adding later that the leader "helps them . . . really identify what are the barriers .. . . 'What's really going on here? And, what can we learn rom this?'"
Support from superiors (H_6)	R: We said, we would give it [teaming] a shot and then we kind of asked [their superiors] "what do you want us to be doing?" They said, "you figure that out." We said, "well, are there any guidelines, what are you looking for?" "You figure that out." . . . A lot of unbelievable work took place while teaming, but we didn't get the direction and the support whether it was office-wise or statewide or whatever the powers that be.	R: I think the management needs to really support the project the way that we said that we were going to do it. INT: And what would that have looked like? R: It would have looked like . . . we would have gotten—maybe not even a hugely reduced caseload—but maybe we wouldn't have gotten new cases and any new cases we got really came in as investigations, as we wanted .... [The director of their office] supports it [teaming], [but] he kind of says . . . "Can't you just team everything?" Which is great but that's not exactly what the proposal said. Our proposal did lay things out pretty clearly. If the things really went the way we said that we wanted it to go . . .. I think it [their teaming] would have been different.	INT: Do you feel like you are getting the resources that you need in order to do your work well and that could be consultation . . . financial resources . . . support from your superiors .. . . R: I think we are relatively lucky in some of those respects because we do have a lot of support especially from [the team's immediate superior.] . . . R: Any time I think we really do need a pep talk, he [their superior] is good to come in, and he can make you feel like you got something when you got nothing. He has that ability.

Continued

Table 1 *(continued)*
Illustrations of Evidence Used to Categorize Teams on Constructs Measured with Qualitative Data

	Low	Medium	High
Experimentation (H_7)	All the teams experimented with multiple new ways of doing things, so we categorized all the teams as High in experimentation. Teams experimented by combining functions, focusing on mentally ill families, regularly bringing in psychologists and other content experts to their meetings, creating meeting rituals that encouraged everyone to participate, providing resources to families immediately after receiving an allegation of abuse or neglect, etc.		
Reflection (H_8)	R: We've been together for three years, and we're still not doing that as a routine process, and I think that's important. INT: Doing what as a routine process? R: Well, reflecting more on the team process and kind of allowing for every member on the team to feel comfortable saying something.	After a member of one team was fired for lying in her paperwork, her team members "all sat down and . . . we tried to process kind of where it started, where we were at, kind of the developments that were made."	R: I remember when we were . . . in that meeting, I felt like we were in therapy all day. You were exhausted. You would never have to go into your inner self and expose yourself to people that way . . . in your regular work environment. It was really good that we were all willing to do that and we did do that. It brought us to the next level."

Note: H, hypothesis; INT, interviewer; R, respondent. Some examples include multiple respondents.
[a]This material comes from an interview with an office director who did not wish her interview to be taped. These quotes are from field notes.

Quantitative Data

As is common in case study research, we did not rely on statistical manipulation of the quantitative data; rather we compared means and then reduced the quantitative data to ordinal categories in order to rank and compare the teams. We used quantitative data to assess six constructs: the team-level variable turnover and the individual-level variables of stress, tenure, education, job satisfaction, and job commitment. We created team means for each of the constructs and used those to rank the teams from one to seven. (One was high, seven was low.) We did this for each of the three rounds, which meant each team had three rankings on each construct. We then arranged the ranks in categories of high, medium, and low.[10]

Then, just as with the qualitative data, we used pattern matching to see if the pattern of the rankings corresponded with the pattern of team learning. As with the qualitative data, we looked to see if change in a given construct over time had an impact on team learning; if not then we collapsed the three rounds into one overall value or categorization of low, middle, or high.

Overall, given the complexity of the data (seven teams, 14 constructs, evaluated at three different points in time), we reduced it to its essence in order to see patterns. What we lost in nuance, we believe we made up in clarity.

Results: Relapse, Retreat, and Re-Creation

In this section, we begin by giving the results on the outcome variable, team learning, as defined by the extent to which the team discarded old routines and created and maintained new ones. Specifically, we track whether they met at least weekly as a team to discuss cases, whether team members informally asked for and provided help to each other, and whether two or more workers were formally assigned to a case. Here we give a brief history of each team, describing the changes they tried to make and whether they were ultimately successful. Following this, we look at the criterion variables, by assessing each hypothesis.

By the time of the third round of data collection—about two and a half years after the teams began their work together—it was clear that the teams varied in the extent of their learning. Two teams had *relapsed*: they had stopped teaming altogether and were working in the traditional way. Two teams had *retreated*: they made significant changes but then lost ground by the third round of data collection. Finally, three had *re-created* themselves by institutionalizing new ways of working. To illustrate these different outcomes, we sketch a brief portrait of the teams in each of these categories.

Relapse to Old Routines

Two teams—Team F and Team G—were unable to establish new ways of working together and ultimately stopped teaming, one after 1 year and one after 2 years.

Team G started with enthusiasm: "We just started teaming everything," said one member. This team carried out assessments, which had to be completed within 45 days and culminated in a service plan for the family. They began meeting weekly to discuss

10 In the first and second round, ranks of 1 and 2 were considered High, ranks of 3, 4, and 5 were considered Medium, and ranks of 6 and 7 were considered Low. In the third round, when there were only five teams still in the study, a rank of 1 was High, ranks of 2, 3, and 4 were medium, and the rank of 5 was Low.

cases and assigned two workers to every case. However, they soon decided that what they were trying was not tenable and, instead, teamed only on selected cases.

However, this did not feel like it was working either and their compliance with paperwork fell behind. Ultimately, they were told by their immediate superior that they had to get their paperwork up to date, whatever the implications for teaming. Eventually, the team stopped teaming, both their weekly meetings and their partnership on cases.

Team F took a somewhat different trajectory but with a similar result. Team members worked only on cases in which at least one of the children had mental health issues. They had weekly meetings at which team members presented cases they were working on. They also established a norm of being able to ask for help. However, although they spent a lot of time discussing how to more formally partner on cases, they did not create a formal division of labor on their casework.

Within a year, they dropped their focus on families with mentally ill children which they had found overwhelming. After 2 years, although they were continuing to meet regularly and help each other informally, they still had not clarified their objectives or formal work processes. At that time, the office leadership disbanded the team altogether, with the members joining other units still working in the traditional way.

Both Teams F and G believed in teaming. "The work that went in with families was unbelievable" said a member of Team G. A member of Team F noted: "There's always a sense of check-in . . . 'Do you need help? How are you doing?'" Yet, despite their best intentions, both groups had relapsed entirely, with members conducting their work using the traditional approach.

Retreat from New Routines

Two teams did make significant changes to their practice in the first year of teaming, but then lost ground, though they did not relapse completely.

Team E had an ambitious mission of combining all three functions, working with families from investigation through assessment and ongoing. They also began meeting once or even twice a week and set a norm of being able to ask for help. They experimented with creating some formal division of labor on a subset of cases, assigning two workers to a case. However, early on, their intention of combining all three functions lagged because the office leaders were not prepared to assign investigation cases to the unit.

By Round 2, the team was faltering. It had new members who were confused about what it meant to be a team and it also experienced some conflict. With little formal division of labor on cases, members asked for help as necessary, but several members said they were more comfortable working on their own. By Round 3, things were worse. Another worker left the team and some of those remaining questioned whether the team was really working differently than any other unit. This team was still meeting weekly and helping out informally, but its attempts to create formal joint responsibility for some of their cases had been lost.

Team D was less ambitious but had a somewhat steadier time of it. They only undertook ongoing work, rather than adding other functions. They did begin meeting weekly and assigning two workers to some of their cases. As of Round 2, their changes were still in place. Members spoke very positively of teaming: "I think it opened the unit more. We know each other better. We communicate more," said a member.

But by Round 3, according to the team's superior, the team had lost ground. "A lot of the innovation that we started with has disappeared . . . Some of the things that [the team

leader] originally wanted to do, he didn't keep up with." Team members were engaged largely in occasional assistance, rather than the formal ongoing division of labor that they had instituted for some families at the beginning. The team had experienced the turnover of two workers and caseloads had been up in the previous year, which team members found undermining.

Although these two teams varied in a number of ways, they did share a similar trajectory. They began by experimenting with several changes: instituting weekly meetings, creating norms which encouraged members to ask for help as needed, and establishing mutual formal responsibility for some of their caseload. By Round 2, Team E was slipping; its formal division of labor on cases was declining. By the third round, Team E had no formal division of labor on cases and Team D's formal partnering had largely disappeared. Both teams had retreated from their routines, though not relapsed completely.

Re-Creation of the Team by Embedding New Routines

Three teams institutionalized new ways of working that remained in place more than 2 years into the Teaming Initiative.

Team C was working only with cases that involved a mentally ill parent, combining assessment and ongoing. They met weekly and encouraged assistance as necessary. They also regularly brought resource people into their team meetings, including several agency staff and someone from another agency specializing in mental health. Unlike any of the other teams, they had no intention of creating a more formal division of labor with two workers assigned to a case because they felt that would not work with their population of mentally ill parents. As of Round 3, their changes were still in place.

At this time, the agency was requiring new changes as part of the "integrated model" it was disseminating to all existing teams. This model required teams to combine investigation, assessment, and ongoing functions and to assign two workers to every case. Team C had four team members trained as investigators, though they had not yet assigned two workers to a case. In short, they had maintained the changes they had carried out 2 years ago and were in the process of making more changes.

Team B made a strong start. They began meeting weekly and encouraged members to ask for assistance as necessary. They combined two functions, assessment and ongoing, and began assigning two workers to some cases. These workers usually stayed together on the case for its duration. The team was seriously rattled when, during the first year, they found that a team member had been documenting work, such as home visits, that had never taken place. At the time of the second team interview, the member had been removed, but the team was still recovering. "It has been trying to get our footing back from losing [that] worker . . . Where we are at today is not even where we were a few months ago," said one member.

However, by Round 3, the team had retained the changes it had made. It was continuing to perform two functions, it was still meeting weekly, and still had formal division of labor on a significant subset of cases. Furthermore, in response to the integrated model now required of all teams, Team B had just assigned two workers to every case. Several workers had just been trained as investigators. Some team members were hesitant about these changes because they felt overwhelmed with their current work, but the team was moving forward.

Team A made the most far-ranging changes of any team and maintained them, more than 2 years on. They established ongoing partnerships on each of its cases, so all 90 or so cases had at least two workers. In addition, Team A was combining all three functions: it

Table 2
Summary of Evidence for Hypotheses Involving Constructs Measured with Qualitative Data

		Re-Created			Retreated		Relapsed	
Hypothesis/ Supported?	Construct	Team A	Team B	Team C	Team D	Team E	Team F	Team G
H_1/yes	Direction	High	High	High	Medium	Medium	Low	Low
H_2/yes	Clarity of composition	High	High	High	Medium	Medium	Low	Medium
H_4/no	Mutual support	High	High	Medium	High	Low	High	High
H_5/yes	Team leader effectiveness	High	High	High	Medium	Medium	Low	Low
H_6/yes	Support from superiors	High	High	High	Medium	Medium	Low	Low
H_7/no	Experimentation	High	High	High	High	High	High	High
H_8/yes	Reflection	High	High	Low	Medium	Medium	Low	Low

started with a case in the investigation stage, carried out the assessment if the case was screened in, and then began the ongoing work with the family. If that work extended beyond 6 months, the team transferred the case to another unit.

As part of this change in function, Team A dramatically changed the timing of services to the families. Traditionally, it could take weeks for families to begin receiving services, after the investigation period and sometimes the assessment. However, Team A began providing support during the investigation phase because they felt immediate support could improve the situation and, perhaps, preclude the agency taking the case. If the case was screened in, it meant a jump start for the family. Team A also experimented with what they called "family forums," a way of bringing multiple stakeholders in a child's life—immediate and extended family, teachers, counselors, and others—together at one time to make decisions about the child's and family's future.

By Round 3, all these changes were still in place, and the team was contemplating further changes: for example, how to include family members as part of their team. Team A was seen as the most effective team; in fact, the agency used Team A's approach as the basis for the integrated model it was requiring of all the other teams.

In sum, although these three "re-created" teams varied in the extent of the changes they made, all had maintained those changes. Across all seven teams, we see great variance in the degree to which they engaged in learning, from those fully able to embed new routines, to those who could do so only partially, to those who could not do so at all.

Assessing the Hypotheses

Here we take each criterion variable and assess the evidence supporting it as an enabler of team learning, as summarized in tables 2 and 3. Overall, we found at least partial support for five of the eight team-level characteristics: clear direction, clear composition, team leader effectiveness, support of superiors, and reflective behaviors. Three others—experimentation, mutual support, and turnover—had no clear pattern of relationship to team learning. Two of the five individual-level factors were found to have some support: job satisfaction and job commitment. Stress level, job tenure, and education level did not appear to have

Table 3
Summary of Evidence for Hypotheses Involving Constructs Measured with Quantitative Data

			Re-Created			Retreated		Relapsed	
Hypothesis/ Supported?	Construct	Measure	Team A	Team B	Team C	Team D	Team E	Team F	Team G
H_3/no	Team turnover	Separation rate[a]	46%	62%	17%	29%	77%	31%	18%
		Rank[b]	3	2	7	5	1	4	6
		Categorization[c]	Medium	High	Low	Medium	High	Medium	Low
H_9/no	Stress	Team mean[d]	3.1	3.2	2.2	3.0	2.5	3.1	3.7
		Rank[b]	3.5	2	7	5	6	3.5	1
		Categorization[c]	Medium	High	Low	Medium	Low	Medium	High
H_{10}/no	Member tenure	In months[e]	115	99	70	99	77	25	133
		Rank[b]	2	3.5	6	3.5	5	7	1
		Categorization[c]	High	Medium	Low	Medium	Medium	Low	High
H_{11}/no	Member education	% w/MA[f]	37%	23%	73%	55%	41%	100%	46%
		Rank[b]	6	7	2	3	5	1	4
		Categorization[c]	Low	Low	High	Medium	Medium	High	Medium
H_{12}/partial	Job commitment	Team mean[d]	3.7	4.0	3.6	3.8	3.9	3.6	3.3
		Rank[b]	4	1	5.5	3	2	5.5	7
		Categorization[c]	Medium	High	Low	Medium	High	Low	Low
H_{12}/partial	Job satisfaction	Team mean[d]	3.8	3.9	3.5	3.6	4.0	3.4	3.3
		Rank[b]	3.5	2	3.5	5	1	6	7
		Categorization[c]	Medium	High	Medium	Medium	High	Low	Low

[a]This figure is the percentage of team members who left over the course of the data collection.
[b]This figure is the team's rank out of seven teams, with 1 as high and 7 as low.
[c]For the first and second rounds of data collection, we categorized ranks 1 and 2 as high, ranks 3, 4, and 5 as medium, and ranks 6 and 7 as low. For Round 3, which included only five teams because two had stopped teaming, we categorized rank 1 as high, ranks 2, 3, and 4 as medium, and rank 5 as low. When teams had the same value on a given variable, we took the average of the ranks involved and gave all those teams the same rank. For example, if two teams both had a mean of 3.1 which would put them at either rank 3 or rank 4, we gave each a rank of 3.5 since that is the average of 3 and 4.
[d]This figure is the mean of the team members, on a scale of 1−5 with 1 as low and 5 as high, across the three rounds of data collection.
[e]This figure is the mean of the team members across the three rounds of data collection.
[f]This figure is the percentage of team members with masters degrees; it is the average across the three rounds of data collection.

an effect. We provide the simple findings in this section, while interpreting and elaborating on these findings in the next section.

H_1—Clear Direction

We hypothesized that the more clarity a team had about its goals and direction, the more successfully it would engage in learning. This was supported as the data show a strong relationship between clear direction and team learning. (Please see table 1 for illustrations of the data we used to categorize teams as high, medium, or low on direction and the other qualitative variables.) Teams A, B, and C, who had all re-created their team practice, had a clear sense of direction. Teams D and E, who retreated from their new routines, had some questions about their goals and purpose. And teams F and G, who relapsed altogether, had no real sense of direction. As a member of Team F said, "I feel like we don't have direction . . . it feels like we're still stumbling at that first stage. It's one thing to work together but what are you working together for? It, sort of, feels like that hasn't been resolved."

Clear direction may be related to having a particular mission to redefine the substance of casework, beyond simply changing the structure of the casework so it involved more than one worker. Teams A, B, and C, all Re-created teams, began their work with the intention of combining casework functions. Team C also had a particular mission of working with families with mentally ill parents. And these three teams all had the clearest direction. Two teams *tried* to change the substance of their casework: Team E (a Retreated team) had hoped to combine all three functions, though they were prevented from doing this by their area office, and Team F (a Relapsed team) had originally worked only with families with a mentally ill child, though they dropped this after about a year because the caseload was too labor intensive. Having lost these initial team goals may have undermined these teams' sense of direction. Two other teams—Team D (Retreated) and Team G (Relapsed)—had no intention to change the substance of their casework, but simply to continue to do what they were doing, with multiple workers. This also may not have been enough of a change to provide the clear focus that enabled full re-creation of the team's work processes.

H_2—Clarity of Team Composition

We hypothesized that teams with greater clarity about the make up of the team would be more likely to succeed at team learning. This was largely supported: Re-created Teams A, B, and C all felt clear on their team membership, whereas Retreated Teams D and E exhibited some confusion over the several-year period. Of the two Relapsed teams, Team F had fundamental questions about their team membership, whereas Team G had had some confusion. The reasons for lack of clarity varied across the four teams.

At the time of the first interview, Retreated Team D was still figuring out who was a member of the team, though this became clearer with time. However, the team often had other employees sitting in temporarily (usually because, with a supervisor leaving, a unit had been disbanded and the employee needed a temporary home) and it was not clear if and how these employees should be included in team meetings and other activities. Retreated Team E began with a clear and expansive sense of their team, including their immediate superior, the director of the office, and an outside psychologist who consulted on cases. However, this became muddier over time, and by the end, the boundaries of the team were not clear to the members.

Team F, one of the Relapsed teams, had significant questions about their membership from the beginning, especially about whether the director of the office was a team member. This was important because she exercised a lot of authority over the team, yet rarely came to meetings, which the team found frustrating. This did not become clearer over time. The only team that does not follow this exact pattern is Team G, the other Relapsed team. Members said they felt clear about its membership, but they also noted that there were two other employees in their same work area, reporting to the same supervisor, who were not team members. That added some ambiguity to the members' sense of membership, though not total confusion.

H_3—Team Turnover

We hypothesized that the more turnover a team experienced, the less likely new routines would be embedded. However, this was not the case. We calculated the "separation rate," which represents the percentage of team members that leave over a given period (Price 1976). Teams A, B, and C, the re-created teams, had separation rates of 46%, 62%, and 17%, respectively. Among the seven teams, these ranked third, second, and seventh. Retreated Teams D and E had rates of 29% and 77%, ranking them fifth and first. And

Relapsed Teams F and G had separation rates of 31% and 18%, ranking them fourth and sixth. Therefore, the two teams with the highest turnover included a Re-created Team and a Retreated team, whereas those with the lowest turnover included both a Re-created team and a Relapsed team.

This lack of relationship is curious and, given the often high turnover rates in child protection services, important. However, we suggest that turnover itself may not be as important as how turnover is managed. Re-created Teams A and B, which both experienced significant turnover, managed the changes by successfully integrating new members. One person from Team A noted, about a time that two members came on simultaneously, "They observed. The first few investigations, they were secondary, and they just kind of sat back." However, the team did meet more often and explicitly discussed "what their [the new members] feelings were—why they are quiet—why they are just taking it in." Further, the team's direct superior met occasionally with the new workers. This informal attention to the transition may have made the difference. Also, these two teams had a clear focus and an effective team leader which could help with integrating workers into a team that was already up and running.

H_4—Mutual Support

We hypothesized that teams with greater mutual trust and support among members would learn more effectively. However, this was not supported. There was relatively little variance among the teams on this score: we assessed five out of seven teams as High in support, meaning that team members felt they trusted each other, enjoyed working together, and believed that others would come to their aid when necessary. Of the two teams not categorized as High, one was a Retreated team and one a Re-created team. A quote from a Team F member illustrates how mutual support could become embedded in the team's work: "The way we help each other. I think we do expect it [help] from each other and expect it because we're a team .... We can read each other's tension." Team E (the Retreated team) stood out among the seven for the amount of intrateam conflict. Unlike any of the other teams, several team members said that they preferred working on their own than asking for help from others. Yet, this team did maintain some of its changes over the duration of the study. Because of the low amount of variation among teams on this variable, it was difficult to determine how mutual support and learning may be related.

H_5—Team Leader Effectiveness

We hypothesized that teams with effective leaders would be more likely to establish new routines, which was supported. The leaders of Teams A, B, and C all had strong assessments; the evaluations were more mixed of the leaders of Teams D and E, and the leaders of teams F and G were seen as the weakest in the sample. The office director of Team A emphasized the importance of the team leader, noting he "really pushes workers to do the best they can, for themselves and for the families," and added later that the leader "helps them ... really identify what are the barriers ... 'What's really going on here? And, what can we learn from this?'"

H_6—Support from Superiors

We hypothesized that teams that felt supported by their superiors would be more likely to learn successfully. This was supported: Teams A, B, and C, who embedded new routines, all felt a strong sense of support from their immediate supervisor and from the director of the area office. Teams D and E, who had mixed success in establishing new practices, felt some support but also expressed some disappointment in support from their superiors. Relapsed

Teams F and G were very critical of their superiors and felt very little support. One team leader pointed to what he felt was the crucial role of his immediate supervisor, "Any time I think we really do need a pep talk, he is good to come in, and he can make you feel like you got something when you got nothing. He has that ability. And he does project that in other people."

H_7—Experimentation

We hypothesized that teams that engaged in more experimentation would be more likely to successfully learn new routines. In fact, this was not supported: there was no variance because all the teams were categorized as High in experimentation.

H_8—Reflection

We hypothesized that teams that engaged in more reflection about their practice would be more likely to learn. This was largely supported as Teams A and B both appeared to engage in reflective practice more often and more deeply than Teams D and E, whereas Teams F and G had almost no evidence of such behavior. The only team that does not follow the pattern is Team C, which we categorized as Low on this construct. This may be because this team made less far-reaching changes in its practice than the other teams—most importantly, they never formally assigned two workers to a case—which perhaps were easier to maintain, even without a reflective capacity.

What does reflection mean in practice? One team described a time when two workers were treated harshly by a family: "They [the family] slammed the door in our face. We had to have the police response out there," said a team member, noting that one worker had been particularly upset by that, whereas the other worker stayed calm and "got us in the door." But rather than singling out the upset worker for blame, the group discussed the incident to see what they all could learn: "We talked about that at group supervision and we really got a lot out of that." When asked what the group learned, the worker continued, "You are not going to be able to handle every case." Another said, "Yes, you are not going to be liked by every client . . . They might be more receptive to your partner." Another added, "Maybe you have to let go . . . or take a back seat on some cases." This team replaced blame on a single worker with a productive conversation for everyone about how to handle the difficulties and complexities of child welfare work.

H_9—Stress

We hypothesized that higher stress would interfere with team learning, but this was not supported. The means across rounds for Teams A, B, and C were 3.1, 3.2, and 2.2, which resulted in rankings of 3.5, 2, and 7 (with higher ranks meaning greater stress.)[11] The overall means for D and E were 3.0 and 2.5, which were fifth and sixth in the rankings. And F and G had overall means of 3.1 and 3.7, ranking 3.5 and 1 in stress. In other words, the two teams that reported the highest level of stress included a Re-created team and a Relapsed team, whereas the lowest levels of stress were reported by a Retreated team and a Re-created team. Therefore, no clear pattern emerged.

H_{10}—Tenure of Team Members

We hypothesized that tenure would undermine team learning but saw no relationship. The members of Re-created teams A, B, and C had an average tenure (across all three rounds of data collection) of 115 months, 99 months, and 70 months, respectively, giving them the

11 When teams had the same mean, we took the average of the ranks involved and gave all those teams the same rank. For example, if two teams both had a mean of 4.0 that would have put them at either rank two or rank three, we gave each a rank of 2.5 since that is the average of two and three.

ranks of 2, 3.5, and 6 among the teams. Retreated teams D and E had an average tenure of 99 and 77 months, 3.5 and 5 among the seven teams. And Relapsed teams F and G had an average tenure of 25 months and 134 months, making them seventh and first in the ranking. In other words, the two teams with the highest average tenure included both a Re-created team and a Relapsed one, and the two teams with the lowest average tenure also included a Re-created team and a Relapsed one.

H_{11}—Educational Level

We hypothesized that teams with a higher percentage of members with master's degrees would be more likely to learn. However, we found no relationship between level of education and extent of learning. Averaged over three rounds, 37% of Team A's members had masters, whereas this was true of 23% of Team B's members and 73% of Team C's—ranking them sixth, seventh, and second, respectively. Teams D and E were at 55% and 41% (third and fifth) and F and G were at 100% and 46% (first and fourth.) Therefore, although the two teams with the least highly educated members were both Re-created, the two teams with the most highly educated members were either Re-created or Relapsed. The three teams in the middle were split between Retreated and Relapsed. Altogether, there is no clear pattern of relationship with team learning.

H_{12}—Job Satisfaction and Job Commitment

We hypothesized that both job satisfaction and job commitment would be associated with higher learning. We found some support for this, in that Relapsed Teams F and G were the two lowest in job satisfaction and were among the three lowest in job commitment. However, the pattern did not hold for the other teams.

We expand on these results in the next section.

DISCUSSION

The ubiquity of routines in street-level practice has been widely recognized (e.g., Maynard-Moody and Musheno 2003; Meyers, Glaser, and Mac Donald 1998). In his seminal work, Lipsky (1980, 149) placed routines at the heart of frontline work: "Street-level bureaucrats . . . develop attachments to modes of practice. They appear to feel that their jobs require the routines." This prompts the question: What enables workers to break old routines and learn new ones?

Previous work on street-level practice has established the importance of individual- and organizational-level factors and some researchers have created models that include both these levels (Jewell and Glaser 2006; Riccucci et al. 2004). However, group-level characteristics have largely been overlooked. Our research provides some evidence that team-level variables matter, with much less support for individual-level factors. We suggest that scholars include the group level in future work on policy, organizational, and street-level change since otherwise we may be overlooking a source of stasis and transformation. Several characteristics seem to play a role.

First, it appears that clear direction may be a key ingredient (cf., Meyers, Riccucci, and Lurie 2001). Clear goals may be important because they provide both the impetus to break routines and a compass that keeps teams on track. Second, clear composition also appears to play a role. This may in part be related to direction, in that it may be difficult to know what you are supposed to accomplish if the cast of characters is changing or unclear. Further, work on the psychology of groups argues that group boundaries are important for a sense of safety and security (Miller and Rice 1967; Yalom 1995).

Third, team leader effectiveness may be important because of the central role leaders play in the other influential factors—setting direction, encouraging reflection, and the like. Fourth, support from the team's superiors could be essential because the team must coordinate with others outside the team. Further, superiors can help the team adjust to—or bend, as necessary—organizational requirements. Fifth, the capacity for reflection would seem to be at the heart of learning: teams must be able to assess their procedures, identify gaps, and consider new approaches.

However, several of our proposed group-level influences did not have an effect. First, we found that experimentation did not distinguish the teams from each other since all were rated High. This seems to be because the Teaming Initiative was itself an experiment and, by definition, all its groups had to try out new ways of working. However, experimentation seems to be a necessary condition for learning (Edmondson 1999), so we expect that research into existing workgroups not engaged in some kind of pilot project would find experimentation to play an important role. It's also possible that experimentation is necessary but not sufficient: groups may try out new ways of working but be unable to make them stick because of a lack of clear mission, or of reflective behaviors, or of effective team leadership.

Second, although turnover varied widely, we found no clear pattern of relationship with team learning. This was unexpected. However, it may be that it is how teams manage their turnover that is important. Re-created Teams A and B had significant turnover, but their clear focus, effective team leadership, and reflective capacity may have helped them integrate workers into a team that was already up and running.

Third, mutual support did not have a clear match with learning, perhaps because there was relatively little variance. However, previous research suggests that trust and connection can lead either to risk taking and innovation (Edmondson 1999) or to insularity and groupthink (Janis 1982; West and Field 1995). What may be more important is whether the teams are in a reflective mode of continuous learning rather than one that seeks comfort and stability.

Although we undertook this research with a particular focus on team-level characteristics, previous research did suggest that individual-level variables could be an alternate explanation. However, we found relatively little evidence that individual characteristics made a difference. We did find some support for the role of job satisfaction and job commitment, in that the two Relapsed teams had low levels of both, though there was no association with any of the other teams. We did not find, contrary to our expectations, that stress was influential; these are stressful jobs, and perhaps, overall, these workers have become conditioned to it. Worker experience, as measured by tenure and educational level, also did not appear to have an effect (cf., Riccucci et al. 2004). Regarding tenure, the selection process for participation largely relied on volunteers (either of full units or individual workers). Therefore, the team members involved, regardless of their years of experience, had an active interest in trying to work in a different way. The story with educational level is less clear, but perhaps the results simply show that an interest in and capacity for learning is not related to formal educational experience.

In sum, this work makes a preliminary case for including team-level factors in future models of what enables learning and change in street-level practice. Further, it may have implications beyond street-level practice since managers and executives often work in groups or teams. Although somewhat different dynamics may be in play, group

characteristics could be influential higher up in the bureaucracy as well. In addition, our work contributes to the literature on team learning in two ways. First, this article stands out because we documented actual changes in routines as a measure of learning as opposed to seeing learning as a process or set of behaviors. As others have argued, defining learning as an outcome, rather than a process, is a more robust and rigorous measure and goes beyond most previous work on team learning (Edmondson 2002). In that sense, this article models one way to operationalize learning as an outcome, which could be useful for other researchers. Second, this article directly compares the impact of team-level variables to individual-level variables which is also relatively rare. The particular constellation of individual characteristics—stress, tenure, education, job satisfaction, and job commitment—has not been explored elsewhere, to our knowledge.

However, this work does have limitations, which narrow generalizability. First, as comparative case study research with a small sample, we can draw only tentative conclusions about the associations that we found. Large-scale research is necessary to confirm these findings. Second, we studied teams of workers who were expected to work collectively and be mutually supportive. Many workgroups operate without those expectations and our findings may not apply to them. Third, we could not investigate organizational-level variables since all the teams were in one organization. Future work should investigate that level as well. Fourth, we wrote this article as a broad brush overview of 13 different criterion variables, which precluded looking more deeply into any of them and how they affected teamwork. Future work will paint a richer, detailed, process-oriented portrait of the teams.

Academics and practitioners alike are interested in how public service agencies can uproot old practices and plant new and adaptive ones. This article adds a new level to our understanding of how that process can happen.

FUNDING

Visiting Scholar Program of the Russell Sage Foundation; the Vladeck Fellowship of New York University; the Massachusetts Department of Children and Families.

REFERENCES

Agranoff, R., and B. A. Radin. 1991. The comparative case study in public administration. *Research in Public Administration* 1:203–31.

Argote, L., D. Gruenfeld, and C. Naquin. 2001. Group learning in organizations. In *Groups at work: Theory and research*, ed. M. E. Turner, 369–411. Mahwah, NJ: Lawrence Erlbaum.

Argyris, C. 1990. *Overcoming organizational defenses: Facilitating organizational learning*. Needham MA: Allyn & Bacon.

Boyatzis, R. E. 1998. *Transforming qualitative information*. Thousand Oaks, CA: Sage.

Boyne, G. A. 2002. Public and private management: What's the difference? *Journal of Management Studies* 39:97–122.

Boyne, G. A., G. Jenkins, and M. Poole. 1999. Human resource management in the public and private sectors. *Public Administration* 77:407–20.

Brewer, G. 2005. In the eye of the storm: Frontline supervisors and federal agency performance. *Journal of Public Administration Research and Theory* 15:505–27.

Buchanan, B., II. 1974. Government managers, business executives and organizational commitment. *Public Administration Review* 34:339–47.

Campion, M. A., G. J. Medsker, and A. C. Higgs. 1993. Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology* 46: 823–50.

Currie, G., R. Finn, and G. Martin. 2007. Spanning boundaries in pursuit of effective knowledge sharing within networks within the NHS. *Journal of Health Organization and Management* 21: 406–17.

Currie, G., and O. Suhomlinova. 2006. The impact of institutional upon knowledge sharing in the UK NHS: The triumph of professional power and the inconsistency of policy. *Public Administration* 84:1–30.

Dekker, S., and D. Hansen. 2004. Learning under pressure: The effects of politicization on organizational learning in public bureaucracies. *Journal of Public Administration Research and Theory* 14: 211–30.

Edmondson, A. 1999. Psychological safety and learning behavior in work teams. *Administrative Science Quarterly* 44:350–83.

Edmondson, A. C. 2002. The local and variegated nature of learning in organizations: A group-level perspective. *Organization Science* 13:128–46.

———. 2003. Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *Journal of Management Studies* 40:1420–52.

Edmondson, A. C., R. M. Bohmer, and G. P. Pisano. 2001. Disrupted routines: Team learning and new technology implementation in hospitals. *Administrative Science Quarterly* 46:685–716.

Edmondson, A. C., J. R. Dillon, and K. S. Roloff. 2007. Three perspectives on team learning: Outcome improvement, task mastery and group process. *Academy of Management Annals* 1:269–314.

Edmondson, A. C., and B. Moingeon. 1998. From organizational learning to the learning organization. *Management Learning* 29:5–20.

Eisenhardt, K. M. 1989. Building theories from case study research. *Academy of Management Review* 14:532–50.

Eisenhardt, K. M., and M. E. Graebner. 2007. Theory-building from cases: Challenges and opportunities. *Academy of Management Journal* 50:25–32.

Ellis, A. P. J., J. R. Hollenbeck, D. R. Ilgen, C. O. L. H. Porter, B. J. West, and H. Moon. 2003. Team learning: Collectively connecting the dots. *Journal of Applied Psychology* 88:821–35.

Fields, D. L. 2002. *Taking the measure of work: A guide to validated scales for organizational research and diagnosis*. Thousand Oaks, CA: Sage.

Freeman, R. 2007. Epistemological bricolage: How practitioners make sense of learning. *Administration and Society* 39:476–96.

George, J. M., and G. R. Jones. 1997. Experiencing work: Values, attitudes and moods. *Human Relations* 50:393–416.

Gersick, C. J. G., and J. R. Hackman. 1990. Habitual routines in task-performing groups. *Organizational Behavior and Human Decision Processes* 47:65–97.

Gibson, C., and F. Vermeulen. 2003. A healthy divide: Subgroups as a stimulus for team learning behavior. *Administrative Science Quarterly* 48:202–39.

Gladstein, D. 1984. A model of task group effectiveness. *Administrative Science Quarterly* 29:499–517.

Goldman, L. S. 2007. Democratic governance across the front-lines of public serving bureaucracies: Structures for learning, collective action and mutual support. Paper presented at the Social Welfare Analysis Colloquium. Boston University, School of Social Work, Boston, MA, November 20, 2006.

Goodman, P. S., E. Ravlin, and M. Schminke. 1987. Understanding groups in organizations. *Research in Organizational Behavior* 9:121–73.

Guzzo, R. A., and M. W. Dickson. 1996. Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology* 47:307–38.

Hackman, J. R. 1987. The design of work teams. In *Handbook of organizational behavior*, ed. J. W. Lorsch, 315–42. Englewood Cliffs, NJ: Prentice-Hall.

Hickman, G. R., and A. Creighton-Zollar. 1998. Diverse self-directed work teams: Developing strategic initiatives for 21st century organization. *Public Personnel Management* 27:187–200.

Hill, H. C. 2003. Understanding implementation: Street-level bureaucrats' resources for reform. *Journal of Public Administration Research and Theory* 13:265–82.
Huber, G. P. 1991. Organizational learning: The contributing processes and literatures. *Organization Science* 2:88–115.
Janis, I. L. 1982. *Groupthink*. Boston, MA: Houghton-Mifflin.
Jewell, C. U., and B. E. Glaser. 2006. Toward a general analytic framework: Organizational settings, policy goals and street-level behavior. *Administration & Society* 38:335–64.
Leisink, P. 2004. Can public management reform be allied with staff participation? *Review of Public Personnel Administration* 24:348–60.
Levitt, B., and J. G. March. 1988. Organizational learning. *Annual Review of Sociology* 14:319–40.
Lewin, K. 1947. Frontiers in group dynamics. *Human Relations* 1:5–41.
Lipsky, M. 1980. *Street-level bureaucracy*. New York, NY: Russell Sage Foundation.
Mahler, J. 1997. Influences of organizational culture on learning in public agency. *Journal of Public Administration Research and Theory* 4:519–40.
Maynard-Moody, S., and M. Musheno. 2003. *Cops, teachers, counselors: Stories from the front lines of public service*. Ann Arbor, MI: University of Michigan Press.
Meyers, M. K., B. Glaser, and K. Mac Donald. 1998. On the front lines of welfare delivery: Are workers implementing policy reforms? *Journal of Policy Analysis & Management* 17:1–22.
Meyers, M. K., N. M. Riccucci, and I. Lurie. 2001. Achieving goal congruence in complex environments: The case of welfare reform. *Journal of Public Administration Research and Theory* 11:165–201.
Miller, E. J., and A. K. Rice. 1967. *Systems of organizations: The control of task and sentient boundaries*. London: Tavistock.
Millward, L. J., and L. J. Hopkins. 1998. Psychological contracts and job commitment. *Journal of Applied Social Psychology* 28:1530–56.
Moynihan, D. P., and S. K. Pandey. 2007. Finding workable levers over work motivation: Comparing job satisfaction, job involvement and organizational commitment. *Administration & Society* 39: 803–32.
Nissly, J. A., M. E. Mor Barak, and A. Levin. 2005. Stress, social support and workers' intention to leave their jobs in public child welfare. *Administration in Social Work* 29:79–100.
Price, J. L. 1976. The measurement of turnover. *Industrial Relations Journal* 6:33–46.
Rashman, L., and J. Hartley. 2002. Leading and learning? Knowledge transfer in the Beacon Council Scheme. *Public Administration* 80:523–42.
Riccucci, N. M. 2005. Street-level bureaucrats and intrastate variation in the implementation of temporary assistance for needy families policies. *Journal of Public Administration Research and Theory* 15:89–111.
Riccucci, N. M., M. K. Meyers, I. Lurie, and J. S. Han. 2004. The implementation of welfare reform policy: The role of public managers in front-line practices. *Public Administration Review* 64:438–48.
Sandfort, J. R. 1999. The structural impediments to human service collaboration: Examining welfare reform at the front lines. *The Social Service Review* 73:314–39.
Scarbrough, H., M. Bresnen, L. F. Edelman, S. Laurent, S. Newell, and J. Swan. 2004. The processes of project-based learning: An exploratory study. *Management Learning* 35:491–506.
Scarbrough, H., J. Swan, S. Laurent, M. Bresnen, L. Edelman, and S. Newell. 2004. Project-based learning and the role of learning boundaries. *Organization Studies* 25:1579–600.
Siggelkow, N. 2007. Persuasian with case studies. *Academy of Management Journal* 50:20–4.
Snyder, W., E. Wenger, and X. de Sousa Briggs. 2003. Communities of practice in government: Leveraging knowledge for performance. *The Public Manager* 32:17–21.
Staw, B. M., L. E. Sandelands, and J. E. Dutton. 1981. Threat-rigidity effects in organizational behavior: A multilevel analysis. *Administrative Science Quarterly* 26:501–24.
Tamuz, M. 2001. Learning disabilities for regulators: The perils of organizational learning in the air transportation industry. *Administration and Society* 33:276–302.
Tsui, A. S., T. D. Egan, and C. A. Reilly, III. 1992. Being different: Relational demography and organizational attachment. *Administrative Science Quarterly* 37:549–79.

Vera, D., and M. Crossan. 2005. Improvisation and innovation performance in teams. *Organization Science* 16:203–24.
Vinzant, J. C., and L. Crothers. 1998. *Street-level leadership: Discretion and legitimacy in front-line public service*. Washington, DC: Georgetown University Press.
Wageman, R., J. R. Hackman, and E. Lehman. 2005. Team diagnostic survey: Development of an instrument. *Journal of Applied Behavioral Science* 41:373–98.
Wenger, E. 2000. Communities of practice and social learning systems. *Organization* 7:225–46.
Wenger, E. C., and W. M. Snyder. 2000. Communities of practice: The organizational frontier. *Harvard Business Review* 78:139–45.
West, M. 1996. Innovation in top management teams. *Journal of Applied Psychology* 81:680–93.
West, M., and R. Field. 1995. Teamwork in primary health care: Perspectives from organizational psychology. *Journal of Interprofessional Care* 9:117–22.
West, M. A., and B. C. Poulton. 1997. A failure of function: Teamwork in primary health care. *Journal of Interprofessional Care* 11:205–16.
Yalom, I. D. 1995. *The theory and practice of group psychotherapy*. New York, NY: Basic Books.
Yin, R. K. 1981. Life histories of innovations: How new practices become routinized. *Public Administration Review* 41:21–8.
———. 2003. *Case study research: Design and methods*. Thousand Oaks, CA: Sage.
Zajac, G., and L. K. Comfort. 1997. "The spirit of watchfulness": Public ethics as organizational learning. *Journal of Public Administration Research and Theory* 7:541–69.