© Academy of Management Journal
2013, Vol. 56, No. 4, 945–971.
http://dx.doi.org/10.5465/amj.2010.0501

CAN SURGICAL TEAMS EVER LEARN? THE ROLE OF COORDINATION, COMPLEXITY, AND TRANSITIVITY IN ACTION TEAM LEARNING

DANA R. VASHDI
The University of Haifa

PETER A. BAMBERGER
Tel Aviv University

MIRIAM EREZ
Technion: Israel Institute of Technology

Recognizing that current theories of team learning do not apply to short-term action teams, we conceptualize how action teams may learn and test hypotheses regarding the performance-related effects of such learning, the mechanisms mediating such effects, and the conditions governing their magnitude. We operationalized the level of action team learning (ATL) in terms of the reglarity and number of role-based, gnided team reflexivity experiences of an action team's members. Testing our hypotheses on 250 surgical teams, we find that higher levels of ATL are associated with shorter surgical duration, with this effect mediated by team helping and workload sharing, particularly under conditions of greater team task complexity. Additionally, we find higher levels of ATL to be directly associated with a reduced number of adverse events in low-complexity surgeries.

Enhancements in knowledge and skills produced by cumulative experience are at the very core of learning, be it at the individual (Weiss, 1990; Wright, 1936), team (Edmondson, 1999; Stagl, Salas, & Day, 2008), or organizational (Argyris & Schön, 1978) level. The mechanisms that enable cumulative experience to generate such competency enhancements are iterative cycles of action and reflection (Edmondson, 1999; Schön, 1983, 1987). In teams, temporal stability—a membership with a "history of working together in the past and an expectation of working together in the future" (Hollenbeck, Beersma, & Schouten, 2012: 84)—provides the foundation for these iterative cycles, and thus for the collective accumulation, storage, and retrieval of shared knowledge (Mohammed & Dumville, 2001; Olivera & Argote, 1999; Rico, Sanchez-Manzanares, Gil, & Gibson, 2008; Wilson, Goodman, & Cronin, 2007). As research on team debriefings and after-event reviews (Blickensderfer, Cannon-Bowers, & Salas, 1997; Salas, Nichols, & Driskell, 2007; Smith-Jentsch, Cannon-Bowers, Tannenbaum, & Salas, 2008) has demonstrated, the shared knowledge generated by such reflection allows team members not only to better anticipate situations, but also to more dynamically adjust to them and to the responses they elicit from other team members—what Rico et al. (2008) refer to as *implicit coordination*. The enhanced team processes facilitated by such learning in stable teams provide a basis for enhanced team performance (Schippers, Den Hartog, Koopman, & Wienk, 2003; Smith-Jentsch et al., 2008; Stasser, Stewart, & Wittenbaum, 1995).

But what if the assumption of team temporal stability is violated? The few studies that have examined this question have consistently found that compositional instability or membership flux seriously impedes team learning processes, and consequently attenuates any association between collective action-and-reflection experiences and team performance (Reagans, Argote, & Brooks, 2005; Van der Vegt, Bunderson, & Kuipers, 2010). Still, al-

All correspondence regarding this article should be addressed to Peter A. Bamberger. The research upon which this article is based was funded by the Israel National Institute for Health Policy Research. The authors wish to acknowledge the assistance of Dr. Ahuva Melnick in data collection. The names of the second and third authors appear in alphabetical order.

though teams with short life spans and transient memberships are an increasingly prevalent aspect of the organizational landscape (Ellis, Bell, Ployhart, Hollenbeck, & Ilgen, 2005; Hollenback et al., 2012), research on team learning has generally focused on teams with stable memberships and life spans covering multiple performance episodes (Chudoba & Watson-Manheim, 2007; Salas et al., 2007). As a result, little is known as to whether, how, and under what conditions team learning may still be possible in short-term or compositionally unstable teams. Indeed, in their review of the team learning literature, Edmondson, Dillon, and Roloff (2008: 308) noted that significant gaps remain in understanding of how learning may occur in teams with "permeable boundaries and transient memberships," and "how one team's learning may affect the team fragments that emerge in later teams."

In an attempt to narrow these gaps, we generate a model of learning in action teams, a classic form of temporally *un*stable (i.e., short-term) teams, and test hypotheses regarding the performance-related effects of such learning, using a sample of surgical teams in a large, tertiary health care center in Israel. Action teams are "highly specialist teams cooperating in brief (but often urgent) performance events that require improvisation in unpredicted circumstances" (Sundstrom, De Meuse, & Futrell, 1990: 12); they include, for example, information technology response teams, surgical teams, and software development teams. Such teams are by definition compositionally unstable, with team members being dispersed to new action teams upon completion of the mission. Although members may have little or no experience working together, these ad hoc teams often face intense, difficult situations that require them to quickly and dynamically respond to multiple task inputs in a highly coordinated manner (Klein, Ziegert, Knight, & Xiao, 2006).

In developing our model of action team learning, we draw from the team training and reflexivity literatures, which suggest that team reflexivity training can enhance the accuracy of team mental models (Marks, Zaccaro, & Mathieu, 2000; Smith-Jentsch et al., 2000, 2008) and "cross-understandings" (Huber & Lewis, 2010). More importantly, this literature suggests that, at least in *stable* teams, team processes and outcomes are likely to be enhanced to the degree that reflective action is structured and guided (Blickensderfer et al., 1997; Smith-Jentsch et al., 2008) to focus on providing team members with the means by which to enhance coordination and nonverbal communication strategies (Salas et al., 2007: 471). We also draw from the suggestion by Edmondson et al. (2008: 308) that such learning processes may have relevance to compositionally unstable teams if one focuses on members' cumulative reflective experiences *across* rather than *within* teams, or in other words, takes into account that such learning processes may be *transitive* in nature. As in the case of multiple team memberships, the transitive nature of such learning stems from members applying lessons learned in prior team experiences to new situations and contexts as they transition from one team to the next (Joshi, Pandey, & Han, 2009; O'Leary, Mortensen, & Woolley, 2011). However, members of multiple long-term teams have the opportunity to spread the shared knowledge gleaned from iterative cycles of action and reflection in each of the teams they belong to. In contrast, the short-term nature of action teams typically precludes such cycles of reflection, thus limiting the *shared* (as opposed to personal) insights (van Ginkel, Tindale, & van Knippenberg, 2009) available for transmission from team to team.

Accordingly, in line with the cumulative experience perspective on learning and recent theorizing on multiple team memberships, we view learning in action teams as a property of a team itself, manifested in terms of the aggregate amount of shared action-reflection experience brought by each member to the team. Moreover, we argue that such learning is likely to yield the most beneficial effects on members' implicit coordination (and, in turn, on team performance) to the extent that (a) members' action-reflection experiences are explicit, shared, and guided to focus on *role holder* behaviors (Gurtner, Tschan, Semmer, & Nägele, 2007) rather than on the behaviors of particular individuals (who, upon mission completion, typically disperse to other teams); (b) these shared action-reflection experiences are numerous, accumulating over multiple action teams and thus expanding the repertoire of behavioral responses available for implicit coordination; and (c) these shared reflexivity experiences occur on a fairly consistent basis, increasing the probability that members will recognize similar patterns of role-based responses to common sets of problems as they move from one team to another. That is, as a *team-level property* we conceptualize ATL as reflecting not simply whether team members are trained in reflexivity methods, but rather the degree to which team members have fairly *consistently* engaged in a *greater* number of *guided,* *shared, and role-focused* reflective experiences following team action.

We posit that higher levels of ATL are likely to be associated with greater implicit coordination (manifested in members' perceptions of peer helping and workload sharing) and, as a result, enhanced team performance. In this context, we define performance with respect to both efficiency-related outcomes (e.g., reduced production cost or time relative to some norm reflecting how well teams leverage knowledge and skills to enhance the efficiency or speed of execution) and quality-related outcomes (e.g., the number or severity of assembly errors or adverse events) (Edmondson et al., 2008; Lewis, Lange, & Gillis, 2005; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000).[1] In addition, because studies on group processes suggest that the impact of such processes on effectiveness is contextually contingent (Gladstein, 1984; Tushman, 1977), we propose that the magnitude of the relationship between ATL and performance is likely to depend on the level of team task complexity.

In addition to drawing from the extant literatures on team training and learning (Gurtner et al., 2007; Salas et al., 2007; Smith-Jentsch et al., 2008), our study offers three main contributions to these literatures. First, by conceptualizing learning in action teams as a team property, we offer a new approach to modeling team learning in compositionally unstable teams, one that takes into consideration *transitive* team learning (i.e., cumulative experience *across* rather than within teams). While others have suggested this notion of learning across rather than within teams (Edmondson et al., 2008), ours is the first study to theoretically explicate and empirically demonstrate how transitive learning may be engineered in those contexts in which more conventional models of team learning are less applicable. Second, we demonstrate that action teams may be well positioned to learn as long as structures are put in place to facilitate—as consistently as possible—members engaging in role-focused shared reflection before they disperse to their next assignments. Previous training studies have demonstrated that by structuring reflexivity in teams, individual (DeRue, Nahrgang, Hollenbeck, & Workman, 2012; Gurtner et al., 2007) and team-level (Blickensderfer et al., 1997; Smith-Jentsch et al., 2008) outcomes may be enhanced. However, our study goes beyond this research in that it suggests that, at least in the context of team temporal instability, what may be most significant is not simply training in structured reflexivity techniques, but the continuous *accumulation* of structured reflexivity experiences. Accordingly, we contribute to team learning theory by specifying and explicating the structural requirements for team learning demanded by temporal instability. Finally, by taking into account implicit coordination as a mediator between ATL and performance, and team task complexity as a context-based moderator (O'Leary et al., 2011), our model responds to Edmondson et al.'s (2008: 305) call for midlevel theories of team learning (i.e., theories explaining how and when relations operate). Such theory is critical if scholars are to begin to understand the mechanisms by which action team learning may affect team performance and the boundary conditions governing the strength of these effects.

A Transitivity-Based Model of Action Team Learning

As noted above, iterative cycles of action and reflection serve as the basis of learning (Mathieu et al., 2000; Schön, 1987). Reflection gives deep meaning to experience (Schön, 1983, 1987) and facilitates behavioral change in response to these inferred meanings at both the individual (Smith, Nolen-Hoeksema, Fredrickson, & Loftus, 2003) and team levels (Argote, Gruenfeld, & Naquin, 2001). Team reflexivity involves a process of "sharing information and reflecting on experience" (Edmondson et al., 2008: 272) in which members collectively examine past behaviors and consider options for change and improvement in team processes and outcomes (Schippers, Den Hartog, & Koopman, 2007; West, 2002). Both process and multiteam membership theories provide important insights into how such reflexivity might serve as the basis for action team learning.

Process theories of team learning suggest that as members engage in multiple episodes of action and reflection, they develop shared understandings of each member's strengths and weaknesses (Edmondson et al., 2008; Mathieu et al., 2000). Addi-

[1] Efficiency-related outcomes parallel the notion of "outcome improvement" capturing the degree to which a team improves its operational efficiency (Edmondson et al., 2007: 273). Quality-related outcomes parallel the notion of "task mastery" capturing the degree to which a team "leverages its members' knowledge and skills to increase the quality and amount of knowledge available for task execution" (Edmondson et al., 2007: 277).

tionally, recent theorizing suggests that with the accumulation of reflexivity experience, teams develop dynamic, moment-by-moment knowledge representations in the form of team situational models (that is, a team-level cognition "associated with a dynamic understanding of the current situation [i.e., environment, task, team] that is developed by team members moment by moment" Rico et al., 2008: 167), as well as a broader set of cross-understandings relating to what team members know, believe in, and are sensitive to (Huber & Lewis, 2010). Among the most critical benefits of such shared knowledge representations and understandings is a heightened level of implicit coordination, which occurs "when team members anticipate the actions and needs of their colleagues and task demands, and dynamically adjust their own behavior accordingly without having to communicate directly with each other or plan the activity" (Rico et al., 2008: 164). And to the extent that such coordination allows a team to develop more effective repertoires of response to dynamic situations, these theories suggest that as members accumulate reflexivity experience over time, team learning is likely to be associated with improved performance (Edmondson et al., 2008; Rico et al., 2008).

Multiteam membership theory suggests that the shared knowledge representations and understandings emerging from action-reflection experience may have relevance and value across teams as well, with cross-team memberships facilitating their transfer from one team to the next (Joshi et al., 2009; O'Leary, Mortenson, & Wooley, 2011). Implicitly, models based on these theories frame team learning as a team property, reflecting not only the shared knowledge and understandings a team generates, but also knowledge imported into the team from other teams with which it is linked via overlapping memberships. These models propose that people who hold multiple team memberships are exposed to a wider variety of shared experiences and thus can bring to a focal team a rich array of process-related insights and innovations (Florin, Lubatkin, & Schulze, 2003). To the extent that knowledge generated by action-reflection cycles in one team is applicable to the uncertainties faced by another, multiteam membership theory suggests that such knowledge may significantly benefit both coordination-related processes and performance-related outcomes for the absorbing team.

With their focus on the potentially transitive nature of team learning (i.e., the notion that teams have the potential to absorb and benefit from process-oriented shared understandings gleaned by members in the context of their membership in other teams), models of multiteam membership provide an important basis for theorizing about action team learning. After all, given the inherent, structural constraints on the development of cumulative experience *within* action teams, ATL is likely to be contingent upon members' ability to bring shared knowledge and understandings with them as they move from one team to the next. However, another assumption of multiteam membership models is that teams have a certain amount of stability, allowing members to engage in the iterative cycles of action and reflection needed to generate knowledge and understandings that might be transferable to other teams or to integrate understandings and knowledge imported from other teams (O'Leary et al., 2011: 470). Indeed, a number of studies indicate that iterative episodes of reflection are critical if a team's members are to recognize and code patterns across tasks and situations—the basis for developing shared understandings of problems and potential solutions (Lewis et al., 2005; Staats, Gino, & Pisano, 2010). Such an assumption poses a considerable but not irreconcilable constraint on any model of ATL, in that action teams generally lack the temporal stability necessary for members to engage in iterative cycles of action and reflection. To overcome such limitations, we propose a model of ATL structured around three main elements.

The first element is the degree to which team reflexivity is (a) structured into the work of action team members and (b) structured to focus more on team-based roles than on individual actors. Regarding the former, research on guided team reflexivity suggests that reflective experiences are more effective when they are built into a team's normal working procedures (Carter & West, 1998). This is so partly for practical reasons—namely, if reflection sessions are not built into a timeline, they are less likely to actually happen, given scheduling and logistic constraints (Haas, 2006). But it also appears that the *intentional* structuring of reflective action into team activities may facilitate the meta-cognition underlying the analysis and enhancement of performance strategies (Gurtner et al., 2007). Regarding the latter, Smith-Jentsch et al. (2008) argued that unstructured team reflexivity can generate insights that are highly person- or situation-specific. Given the compositional instability of action teams, such person-situation specificity could greatly reduce the transferability of new insights as members move from one team to the next. Accordingly, to the extent that team reflexivity is structured to focus on the needs, repertoires, and situational responses of the various *roles* that members encounter as they move from team to team, pattern recognition and the development of more generalizable, shared understandings may be facilitated.

The second element is the size of the portfolio of reflexivity experiences that members bring with them to an action team. As noted above, research on team reflexivity suggests that the recognition of patterns and routines underlying the development of shared understandings typically requires *multiple* action-reflection experiences (Lewis et al., 2005; Staats et al., 2010). Indeed, Gurtner et al. (2007), in a lab-based simulation using structured reflection, found that a one-time reflexivity episode failed to yield a significant performance improvement for experimental versus control groups. Given that action teams dissolve upon the conclusion of their mission, action team members must accumulate their portfolios of action-reflection experiences *across* rather than within action teams. Accordingly, to the extent that these reflective experiences are role-based (making insights gleaned from them generalizable across action teams), a greater number of such experiences among action team members is likely to provide a broader and more varied basis for pattern recognition and the development of shared understandings.

Third, the *regularity* of action-reflection experiences is likely to matter as much as their quantity. Recent research suggests that individuals' ability to recognize links between prior experience and current stimuli is influenced by the frequency with which active learning occurs (Erev & Haruvy, 2010; Hertwig & Erev, 2009). Contemporary research into the physiology of the brain supports this contention. For instance, Kandel's (2007) work on long-term potentiation in the brain shows that proteins have to be synthesized to convert short-term memories into long-term ones, and this is most likely to occur following high-frequency stimulation of chemical synapses. It thus appears that team members who engage in continuous rather than scattered reflexive experiences will be better able to recognize varying patterns of role-based actions and responses and to anticipate and adjust to the actions of a continuing stream of new and often unfamiliar teammates as they move from one action team to the next. Thus, to the extent that more regular experiences characterize team members' portfolios of guided, role-based reflexivity, the experiences are likely to enable a team to build upon a richer set of shared insights and understandings, thus further contributing to the team's processes and outcomes.

We propose that all three elements of our model are necessary for significant action team learning to take place. In other words, any one or two elements independently—structured reflection, a greater number of reflection sessions, or greater regularity of reflection sessions—may contribute somewhat to ATL, but measurable learning in action teams requires all three. Moreover, from the process models of team learning noted above (and described in greater detail below), we expect that in action teams with greater ATL, a larger and richer body of shared insights and understandings should shorten task completion times (an important indicator of efficiency-related outcomes [Edmondson et al., 2008]) and reduce the occurrence of adverse events (a key indicator of quality-related outcomes [Edmondson et al., 2008]). Accordingly:

Hypothesis 1a. There is an inverse relationship between an action team's level of action team learning (ATL) and the number of adverse events experienced during the action team's mission.

Hypothesis 1b. There is an inverse relationship between an action team's level of ATL and the amount of time relative to the norm it needs to complete its mission.

The Mediating Role of Task-Related Team Processes

As noted above, recent research suggests that enhanced team processes, and particularly those that are coordination-related, link team learning to performance outcomes. For example, both Rico et al. (2007) and Huber and Lewis (2010) developed models suggesting that the association between shared experience and understanding on the one hand, and team performance on the other, is mediated by implicit coordination. And in one of the few empirical studies examining how members' shared understandings affect team performance, Mathieu et al. (2000) demonstrated the key mediating role played by effective team processes, particularly those related to *explicit* coordination.

Although we are unaware of any empirical research examining how the *implicit* coordination elicited by ATL may manifest itself in teams, Rico et al.'s (2008) conceptual model suggests two coordination-related processes, namely workload sharing and helping, as key mechanisms linking ATL to performance-related outcomes in action teams. More specifically, Rico et al. argued that implicit coordination is characterized by "proactively sharing a workload or helping a colleague" (2008: 165), among other things. Workload sharing and helping are distinct, in that workload sharing reflects the degree to which team members reshape their in-role behaviors to more fairly and effectively allocate the team's work among its members (Erez, LePine, & Elms, 2002), while helping reflects extra-role behavior typically provided in response to a solicitation for such assistance (Organ, Podsakoff, & MacKenzie, 2006). Yet the two processes are similar, in that both require team members to anticipate the task demands, actions and needs of fellow team members, and dynamically adapt their behavior to them (Rico et al., 2008). Given that previous research has also demonstrated the positive impact of these two coordination-related mechanisms on team performance (Erez et al., 2002; Ng & Van Dyne, 2005) our theorizing focuses on how ATL influences both mechanisms.

The mediating role of workload sharing. Workload sharing—the effective and equitable allocation of team tasks (Erez et al., 2002)—represents an important form of implicit coordination in work teams. Because it reflects team members' ability to understand one another's role demands and their interrole interactions "without the need for overt communication" (Rico et al., 2008: 165), it is likely to be enhanced as a function of members' experience working with one another and impeded as a function of team compositional instability. However, ATL may provide the basis for workload sharing even in compositionally unstable teams, in that members of teams characterized by greater ATL are likely to have (a) reflected upon role-specific areas of expertise and the optimal distribution of team tasks across roles and (b) developed shared, dynamic understandings of behavioral repertoires appropriate to different role holders, allowing them to recognize and anticipate each other's needs and actions as situations unfold. With a more developed understanding of the needs and skills of each role holder and of how a team's workload may need to be dynamically redistributed as situations change, tasks are likely to be more effectively and efficiently shared. Thus, for example, although the roles of surgeon and nurse in a surgical team are clear and unambiguous, unpredictable events may require some degree of role fluidity. If the surgeon is focused on trying to manage unexpected bleeding, it is up to the nurse to keep an eye on the heart rate monitor. Following this reasoning, ATL can help team members more effectively leverage their knowledge, skills, and attention (Edmondson et al., 2008) to improve workload sharing, and thereby to improve performance, as measured by the occurrence of adverse events and mission completion time (Fiore, Salas, & Cannon-Bowers, 2001). Accordingly, we posit:

Hypothesis 2a. Team workload sharing mediates the relationship between ATL and the number of adverse events experienced during a team's mission.

Hypothesis 2b. Team workload sharing mediates the relationship between ATL and the amount of time—relative to the norm—needed for a team to complete its mission.

The mediating role of team helping. Helping behavior serves as the second pathway by which the implicit coordination elicited by ATL may enhance the performance of action teams. Research has consistently shown helping to have a beneficial impact on team performance (Podsakoff, Whiting, Podsakoff, & Blume, 2009; Spitzmuller, Van Dyne, & Ilies, 2008).

We propose that teams characterized by higher levels of ATL are likely to demonstrate a greater degree of team helping, in that their members are likely to be more attuned to deviations from expected patterns of behavior that may indicate that a role holder is having difficulty and requires help. Accordingly, members of such teams may be more proactive in helping, rather than waiting—often in vain (Bamberger, 2009)—for those in need of assistance to acknowledge the situation and seek help. Similarly, these same shared experiences may make potential help providers more aware of the benefits to the team of providing help, and thus more prepared to accede to even the most implicit of help solicitations. Hence:

Hypothesis 2c. Team helping mediates the relationship between ATL and the number of adverse events experienced during a mission.

Hypothesis 2d. Team helping behavior mediates the relationship between ATL and the amount of time—relative to the norm—for a team to complete its mission.

The Moderating Role of Team Task Complexity

Tasks are more complex when they require an unpredictable number of distinct steps and involve the processing of multiple informational cues (Wood, 1986). Team task complexity consequently reflects the degree to which tasks are defined, structured, and predictable and thus easily managed by means of standardized procedures (Weingart, 1992; Xiao, Hunter, Mackenzie, Jefferies, & the Lotas Group, 1996).

Team task complexity and the indirect effect of ATL on team performance. Drawing on the contingency models of organizational behavior developed in the 1960s (Lawrence & Lorsch, 1967; Thompson, 1967), scholars have long argued that the impact of team processes on team effectiveness is contingent upon task-related conditions, with the effects amplified when team tasks are more complex and attenuated when tasks are simpler (Gladstein, 1984; Stewart & Barrick, 2000). For example, Gladstein proposed that "those process variables that increase (a group's) information-processing capacity will be more predictive of group effectiveness with complex tasks . . . than they will with tasks that are simple" (1984: 501). Similarly, Stewart and Barrick argued that intrateam processes are most predictive of team performance when the "ends and means of production are unclear, requiring team members to interact in novel ways to determine how to proceed" (2000: 137). This logic suggests that, particularly in the context of more complex team tasks, the implicit coordination unleashed by ATL allows team members to more effectively and efficiently leverage their knowledge and skills to accomplish their mission more quickly, and with a smaller number of adverse events.

Team task complexity is likely to amplify the link between implicit coordination and quality-related outcomes (i.e., reducing the number of adverse events) with regard to both workload sharing and helping. While greater workload sharing and helping may offer some utility in straightforward team tasks, they are likely to have greater impact when tasks are unusual and standard operating procedures are less applicable, requiring team members to reallocate their workload in unpredictable ways. As Horwitz and Horwitz noted, team members must "pull together their diverse expertise to formulate strategies to deal with tasks under complex conditions" (2007: 995). These same authors suggested that some forms of helping may not only have less beneficial effects under conditions of low complexity, but, by drawing attention away from routine tasks, may even be counterproductive. Accordingly, we posit:

Hypothesis 3a. Task complexity amplifies the relationship between workload sharing and adverse event occurrence: The beneficial effects of workload sharing on reducing adverse events increases as a function of increased task complexity.

Hypothesis 3b. Task complexity amplifies the relationship between helping and the occurrence of adverse events: The beneficial effect of helping on reducing adverse events increases as a function of increased task complexity.

For similar reasons, greater team task complexity is also likely to amplify the impact of both workload sharing and helping on efficiency-related outcomes (i.e., mission duration). Standard operating procedures for low-complexity projects typically take efficiency considerations into account, and in many cases they are even structured around principles of efficiency (Adler, Goldoftas, & Levine, 1999; Thompson, 1967). This reduces the likelihood that greater workload sharing or helping will offer any meaningful reduction in mission duration. Indeed, because even implicit workload sharing and helping demand the allocation of temporal resources, in low-complexity situations the time required to reallocate the work or provide assistance may ultimately be greater than the time potentially saved. As team tasks become more complex, standardized protocols become less applicable. Teams handling complex projects will work more efficiently if they are prepared to reallocate tasks in response to situational contingencies and if team members are attuned to teammates' need for help. Greater team task complexity thus increases the importance of shared understandings with regard to how the workload should be distributed or when and how help should be offered (Stewart & Barrick, 2000). Accordingly, we posit:

Hypothesis 3c. Task complexity amplifies the relationship between workload sharing and relative task duration: The beneficial effect of workload sharing on reducing task duration increases as a function of increased task complexity.

Hypothesis 3d. Task complexity amplifies the relationship between helping and relative task *duration. The beneficial effect of helping on reducing task duration increases as a function of increased task complexity.*

Team task complexity and the direct effect of ATL on team performance. Just as we posit that team task complexity moderates the indirect effect of ATL on team performance (via workload sharing and helping), we also posit that team task complexity is likely to moderate the unmediated, *direct* effects of ATL on both adverse event occurrence and mission duration. These direct effects likely reflect domain (i.e., content-based) knowledge derived from prior team reflexivity experiences (Smith-Jentsch et al., 2008). After all, as others have suggested (O'Leary et al., 2011; Wilson et al., 2007), team learning involves not only improving team processes, but also enhancements in the domain knowledge held by a team's members. And while the literature on team processes suggests that the association between team processes and performance is *amplified* under conditions of greater task complexity, the literature on multiteam memberships (O'Leary et al., 2011) suggests that the impact of prior team domain knowledge on focal team performance is likely to be *attenuated* as a function of task complexity. Underlying this argument is the notion of analogical dissimilarity, or the idea that as complexity increases, the relevance of domain knowledge gleaned from prior team reflexivity experiences becomes more limited.

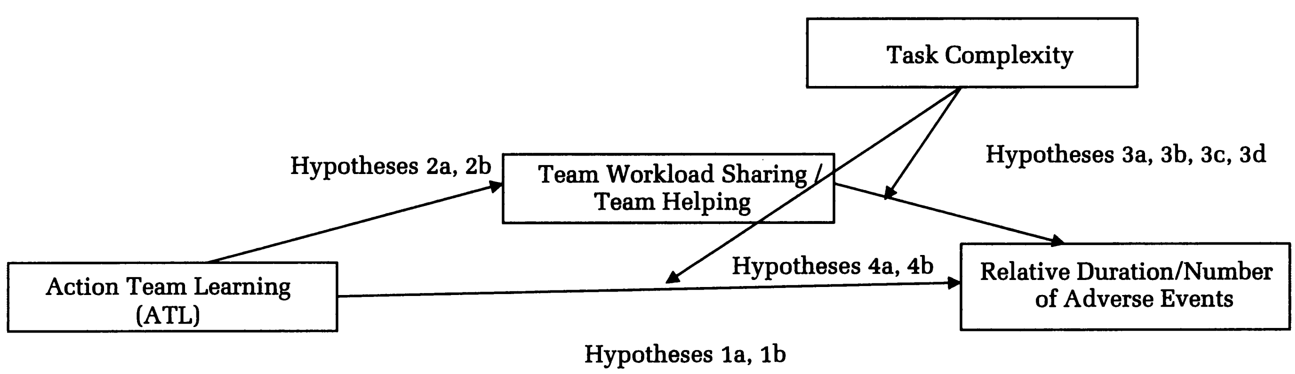
O'Leary et al. (2011) argued that a greater variety of prior experiences among team members is associated with diminishing returns in terms of team learning, in that experiential diversity complicates the interpretation and integration of domain knowledge. This suggests that shared understandings of means-ends relations in action teams are likely to be based more on members' reflective analysis of situations and experiences that are common and generalizable across teams than on unique cases. Further, even when team members fill the same role across teams, it may be difficult to apply general understandings gleaned from prior reflexivity experiences to more complex situations. In particular, attention fragmentation may occur as team members attempt to solve problems according to poorly fitting understandings, in that they will likely need to split their focus between a task itself and the need to adapt these understandings to new contingencies. These effects on attention are likely to increase the risk of adverse events (Kostopoulou & Delaney, 2007). Additionally, the time needed to adapt ATL-generated shared understandings to new contingencies (not to mention the time needed to correct any problems resulting from poor fit) is likely to offset much of the time saved by the application of these shared understandings in the first place (Huey & Wickens, 1993; O'Leary et al., 2011). Accordingly, we posit:

Hypothesis 4a. Team task complexity attenuates the direct relationship between ATL and the number of adverse events: The beneficial effects of ATL are greater at lower levels of task complexity.

Hypothesis 4b. Team task complexity attenuates the direct relationship between ATL and the amount of time—relative to the norm—needed for a team to complete its mission: The beneficial effects of ATL are greater at lower levels of task complexity.

The full model is shown in Figure 1.

FIGURE 1
Proposed Model



METHODS

Data and Sample

We tested our hypotheses using a sample of surgical teams in a large, public, tertiary health care center in Israel. Three hundred sixty-two surgical teams nested in nine surgical wards participated in the study over six months. Each team had three to eight members and included at least one surgeon, one nurse, and one anesthesiologist. Owing to the large pool of surgical and nursing staff and their schedule variations, fewer than 5 percent of the teams shared the exact same composition.

To ensure variance in our primary independent variable, action team learning (ATL), we purposely manipulated the degree to which members of particular wards would have the potential to participate in guided team reflexivity. Management was asked to identify sets of three wards that were similar in terms of their particular medical/surgical specialty and scope of surgical activity, resulting in three sets of three matched wards. To reduce the risk of confounding effects, we randomly selected one ward from each matched set for the manipulation, with the other two constituting controls. In each ward in the intervention condition, ward chiefs agreed to have their staff trained in guided reflexivity and to subsequently try to implement and integrate such practices (including pre- and postaction debriefings) as part of the ward's daily activity.

Design

The study was designed as a longitudinal field study with three phases.

Baseline phase. To ensure that there would be no systematic differences with regard to the performance outcomes (duration and occurrence of adverse events) among the different wards, preintervention performance data were collected from 112 surgical teams in all nine wards.

Training phase. Once the baseline data were collected, surgical team members in the three wards assigned to the intervention condition underwent training in guided reflexivity on the basis of a briefing-debriefing model drawn from the Israeli Air Force (Ellis & Davidi, 2005; Ron, Lipshitz, & Popper, 2006; Vashdi, Bamberger, Erez, & Weiss-Meilik, 2007). The hospital's administration approved two guided reflexivity protocols that we developed—one each for the briefing and debriefing. The briefing protocol covered (a) the indications leading to an operation, (b) the procedure to be performed, (c) the kind of anesthetic to be used, (d) special equipment needed, (e) possible complications, and (f) protocols to be followed in the event such complications arose. The debriefing protocol included a review and analysis of (a) what happened during the surgery, (b) any problems or complications that arose, (c) the degree to which surgical goals were met, (d) what prevented the achievement of specific goals, and (e) what might be done in the future to avoid such complications and to better assure the meeting of objectives. Staff members from the three intervention wards were trained in how to follow these two protocols, with one of the surgeons (either the head surgeon or an assisting surgeon) leading the process. Surgical nurses, anesthesiologists, and other support staff (such as heart-lung technicians) typically assigned to work with the three intervention wards underwent the training together with the surgeons from these wards.

Performance phase. Immediately upon conclusion of the training, staff in the three intervention wards began to implement guided reflexivity practices. Surgeries were randomly selected for observation over a four-month period. Of the surgeries observed, 135 were performed by teams in which the surgeons belonged to one of the intervention wards, and an additional random sample of 115 surgeries were performed by teams in which the surgeons belonged to one of the six control wards.

Data were collected by senior medical students, all of whom received a full day of training in observing surgeries and using our observation protocol. For each randomly selected surgery, the trained observers were asked to record whether a team conducted a full briefing-debriefing process, a partial process (either a briefing or a debriefing), or none at all. The observers were completely blind to the experimental conditions. The observers, present throughout each entire surgery, also documented performance outcomes (duration and adverse events) for all surgeries, regardless of the experimental condition, and completed a form listing all the surgical parameters covered by our measures. These parameters, as well as the procedures used for collecting and recording these observations, are discussed below.

Of the 135 teams observed in the intervention condition, 59 implemented both a briefing and a debriefing. Four teams conducted only a debriefing, and 17 conducted only a briefing. The remaining 55 teams implemented no briefing or debriefing at all. While the heads of the three wards encouraged their staff to perform the full briefing-debriefing for every surgery, the teams that failed to implement the full process cited various reasons, primarily time pressure and difficulty rounding up the team's members either right before or right after the surgery. There was no difference found in the mean level of complexity between the surgeries performed by those that conducted briefing-debriefings and those that did not (mean $\text{complexity}_{\text{briefing-debriefing}}$ = 1.83, mean $\text{complexity}_{\text{no briefing-debriefing}}$ = 2, t_{120} = 1.05, n.s.).

Measures

Action team learning (*ATL*) was operationalized as a team composition variable by assessing the aggregated level of *continuously accumulated* team reflexivity experience of a team's members. More specifically, we measured ATL on the basis of the regularity and amount of members' postaction debriefings. In all but four of the 63 teams that conducted a debriefing, a preaction briefing was also performed. Observers confirmed that in all 63 cases, the debriefing was characterized by sharing, discussion, and analysis of team experiences and that team members followed the written guidelines of what should be discussed (and how) in such debriefing sessions.

To generate our measure of ATL, we calculated for each team member (a) the number of sampled surgeries in which he/she participated and (b) the number of these surgeries in which a debriefing was conducted. We then aggregated these numbers to the team level, assessing a team's total reflexivity experience by calculating the total number of sampled surgeries in which a debriefing was conducted involving one or more of the team's members. To calculate the level of action team learning for each team, we then weighted this measure by the regularity of such participation. The latter was calculated as the number of sampled surgeries involving the team's members in which a debriefing was conducted as a proportion of the total number of sampled surgeries in which the team's members participated. Accordingly, ATL was calculated as the product of amount and regularity and ranged from 0 to 24. It is important to note that as ATL is a team-level construct and as nurses and anesthesiologists conduct surgeries with surgeons from different wards, ATL may be more than zero even in teams in which the surgeons did not belong to wards assigned to the intervention condition. For example, a team led by a surgeon from a control ward might have an ATL greater than zero if one or more nurses on that team had previously been part of a team led by a surgeon from one of the intervention wards who conducted a reflexivity session.

Number of adverse events. Following Gawande, Thomas, Zinner, and Brennan (1999: 67), we defined an adverse event as any unusual or irregular event occurring during any phase of surgery that resulted from surgical management and not disease and that posed a potential or actual risk to the patient. Using this conceptualization, Gawande et al. (1999) found life-threatening adverse events to have been recorded in just 3 percent of the more than 14,000 surgeries they sampled. However, as Gawande et al. (1999: 73) noted, because "record review captures only adverse events documented in patient records," a focus on written medical records alone is likely to "lead to an underestimate of adverse events." Indeed, many adverse events are not recorded because, although they may have had the potential to be harmful, they did not result in any damage. In addition, because of concerns about legal liability, what is written in medical records is typically very limited.

Consequently, rather than relying on patient records, under our protocol, observers used a checklist to record all adverse events, whether life-threatening or not. Items on the checklist included, for example, "Was there a match in the counts of surgical pads before and after the surgery?" The first author developed the checklist with the assistance of hospital quality assurance professionals and medical staff from all nine of the wards studied. Some items on the list were drawn from the research literature (e.g., Gawande et al., 1999; Roth et al., 2004), and some from archival sources (e.g., a review of hospital and Israeli Ministry of Health memoranda and protocols regarding surgical regulations and preventable errors). We assessed the reliability of this instrument through an independent pilot sample of 15 operations, each observed by two medical students. We calculated reliability as the degree of agreement between the two observers in each pair on all the checklist parameters, with each parameter coded as 1 if the observers completely agreed and 0 otherwise. Among the 15 sets of observations analyzed, interrater reliability ranged between 75 and 100 percent, and averaged 86 percent. Additionally, as our data are ordered, we also estimated agreement between the two observers on the basis of a quadratic-weighted kappa statistic (Fleiss, Levin, & Paik, 2004). Kappa was .67, which, according to Altman (1991), indicates good agreement.

As in Gawande et al.'s (1999) findings, the distribution of adverse events in our study was skewed. Observers recorded three adverse events in 4 percent of the surgeries, two events in 9 percent, one in 30 percent, and none in 57 percent of the surgeries. The variable was thus defined as the number of adverse events that occurred during a specific surgery, with a range of 0–3.

Relative duration. Our observers recorded the time each patient entered the operating room and the time the operation was declared over. Because surgeries vary widely in length across subdisciplines (ranging from ten minutes to 10.5 hours), we normalized this variable by calculating the ratio of the duration of the particular surgery observed to the mean duration of the same types of surgeries performed by teams in the same ward during the baseline period.

Team workload sharing. Using the scale developed by Erez et al. (2002), participants rated (1 = "strongly disagree" to 7 = "strongly agree"; $\alpha = .86$) four statements describing behaviors of their team relevant to workload sharing (e.g., "Each team member does his/her share of the work on the team's task"). We aggregated the responses to the team level by taking the mean score of the members' responses. The mean coefficient of agreement (r_{wg}) for all teams was 0.82, with scores ranging from 0.4 to 1(ICC1 = 0.13, ICC2 = 0.25). Although ICC2 was below conventional levels, Bliese noted that particularly when influenced by small group size, "lower values may be justified if, despite relatively unreliable group means, one still detects emergent effects" (personal correspondence, 2006). It should be noted that these data were collected at the end of each surgery but prior to the debriefings in cases where these took place.

Team helping. We conceptualized team helping as team members' shared perceptions of the degree to which members provided assistance to fellow team members. Accordingly, we assessed helping from the perspective of both the giver and the recipient. Following the work of Anderson and Williams (1996), at the end of a surgery (but prior to any debriefing), we gave each participant a list of team members and the following instructions: "During surgery, team members are likely to encounter professional problems involving planning or implementation of one's role. In order to deal with these problems, you can turn to any of your team members for help in order to get another perspective on the matter, information, guidance or actual help. Thinking about the task-related problems you confronted in the surgery just completed, and using the attached list of team members you just worked with, please indicate the extent to which you *received* help from each of these individuals. Please [also] indicate the extent to which you *provided* task-related assistance to each of those on the list in the surgery just completed." Participants responded using a scale ranging from 1, "very little," to 7, "very much." To gauge the level of helping within teams, we first averaged self-assessments of help provided by each team member. For example, if A assessed the help he/she provided to team members B, C, and D at levels 2, 3, and 4 respectively, A's help giving was coded as 3. We similarly assessed the average help received at the individual level. To invert the example above, if A received help at levels 2, 3, and 4 from members B, C, and D respectively, A's help receiving was coded as 3. We then calculated the shared perception of team helping for each team as the average of these individual-level mean scores. The two measures of helping were highly correlated ($r = .74$), thus further justifying their combination. The mean coefficient of agreement (r_{wg}) for all teams was 0.72, with scores ranging from 0.15 to 1 (ICC1 = 0.10, ICC2 = 0.19; $\alpha = .92$).

As the measures of team workload sharing and helping were collected in the same questionnaire, we conducted a confirmatory factor analysis to ensure that we were examining two distinct variables. The results supported a two-factor model. Specifically, the two-factor model produced acceptable fit indexes ($\chi^2_{44} = 78.02$, $p < .01$, CFI = 0.97, NFI = 0.94, RMR = 0.10, and RMSEA = 0.08) and fit significantly better than the alternative one-factor model ($\chi^2_{45} = 367.22$, $p < .0001$, CFI = 0.72, NFI = 0.70, RMR = 0.62, and RMSEA = 0.24; $\Delta\chi^2 = 289.2$, $\Delta df = 1$, $p < .001$).

Team task complexity. Surgical tasks, relative to most other occupational tasks, are all highly complex (Wood, 1986). Nevertheless, the degree of task complexity encountered by surgical teams can hardly be deemed invariant. Ethnographic studies (Hazelhurst, Mcmullen, & Gorman, 2004; Pope, 2002) indicate that surgeries vary in complexity, with more complex surgeries characterized by a greater number of procedures that must be performed in an integrated fashion (i.e., either simultaneously or in rapid succession), a need for more intensive verbal and nonverbal coordination among surgical staff, and greater overall uncertainty regarding what needs to be done (i.e., patients must often be "opened up" before a treatment plan can be determined) or how patients will respond. We assessed team task complexity in *relativistic* terms as perceived by the team's head surgeon. At the close of each operation, but prior to any debriefing, the observer asked the head surgeon, "Relative to the surgeries typically performed by you and your colleagues in your ward, how complex would you rate the surgery that you and your team just completed?" (1 = "less complex," 2 = "average," and 3 = "more complex"). Single-item scales have been used effectively for the assessment of a wide variety of organizational constructs, including job satisfaction (Wanous, Reichers, & Hudy, 1997), social identity (Bergami & Bagozzi, 2000), and social identification (Shamir & Kark, 2004). To assess the interrater reliability of this measure, we asked the head surgeon and one other member of each surgical team in the independent pilot sample of 15 operations to respond to this question independently. Interrater agreement was 100 percent. The Appendix presents additional evidence of construct validity.

Control variables. To rule out possible confounding effects, we controlled for a number of patient and team characteristics. First, to control for possible effects of patients' general health, we recorded each patient's ASA score—a metric of preoperative physical status developed by the American Society of Anesthetists. In a *Annals of Surgery* article, the ASA was found "to be one of the top 10 predictors (out of a possible 60 preoperative risk factors) for morbidity and mortality in 8 separate surgical specialty models" (Devenport, Bowe, Henderson, Khuri, & Mentzer, 2006: 639). With higher levels of ASA indicative of higher risk, in a given surgery it is likely to be related with increased precautions taken during the surgery (which may lengthen surgical duration) as well as a higher likelihood of adverse events. ASA categories range from 1 (normal, healthy patient) to 5 (moribund patient who is not expected to survive another 24 hours with or without surgery). ASA levels in the current study ranged from 1 through 4. Second, to demonstrate the effect of ATL over and above any premission briefing, we controlled for the occurrence of preoperative briefings. This variable was coded as 0 if no preoperative briefing was conducted and 1 if a briefing occurred. Because data on the dependent and mediating variables were collected prior to the postaction review, there was no need to control for debriefings. Third, we controlled for team size (i.e., the number of team members taking part in a given surgery). Additionally, to show that the performance effects we ascribe to ATL are independent of members' experience working together, we controlled for the number of prior surgeries including at least two members of a given team. Given that this number might be impacted by the size of the ward in question (smaller wards are likely to have a smaller pool of staff to draw from), we centered this variable by subtracting a value equivalent to the mean number of surgeries in which at least two team members had jointly participated. Thus, this variable represents the degree to which shared experience by at least two team members deviates from the mean for that ward. Finally, variance in surgical team performance may also stem from differences in experience, surgical style, or team leadership behaviors exhibited by the head surgeon. For example, through more effective leadership, some surgeons may be able to complete procedures more quickly, encourage safe practices, and encourage reflection. To control for such effects, we incorporated into our models a parameter accounting for the possibility of such random variance (described below).

Data Analysis

As the data were collected from teams in nine different wards, our analysis employed random coefficient modeling (RCM; Goldstein, 1987). This approach allows for testing the nesting of surgeries by head surgeons and the nesting of head surgeons within surgical wards. The advantage of RCM is that by modeling residuals at levels 2 and 3 (with the surgery serving as the level 1 unit of analysis), such analysis acknowledges that surgeries conducted by the same head surgeon and/or performed within the context of the same ward may be more similar to one another than to surgeries conducted by different head surgeons and/or performed by surgical teams affiliated with different wards (Bryk & Raudenbush, 1992). We analyzed our data using the SAS-MIXED procedure when the dependent variable was continuous (i.e., relative duration of surgery) and the SAS-GLIMMIX procedure indicating a negative binomial distribution when the dependent variable was a count variable (i.e., number of adverse events). In all models specifying (head surgeon–assessed) task complexity as a moderating variable, we included a random effect for surgeon-perceived task complexity to account for any surgeon-based (i.e., within-person) variance.

We framed our analysis around the moderated-mediation model implied by our hypotheses. As Muller, Judd, and Yzerbyt noted, "moderated mediation happens if the mediating process that is responsible for producing the effect of the treatment on the outcome depends on the value of the moderator variable. . . . If the moderator is a contextual variable, then it would mean that the mediating processes varies as a function of the context" (2005: 854). Accordingly, we first tested the main effect of ATL on the two performance outcomes (Hypotheses 1a and 1b). Following this, we assessed the mediating effects of team workload sharing and team helping specified in Hypotheses 2a and 2b, again taking into account the nested structure of the data. Finally, we tested the mediated and direct paths specified in Hypotheses 3a, 3b, 3c, 3d, 4a, and 4b in the context of a broader model based on Edwards and Lambert's (2007: 10) equations for a model incorporating second-stage and direct effect moderation. We then estimated the sampling distribution of the indirect effects nonparametrically through bootstrapping and used information from the bootstrap sampling distribution to generate confidence intervals for the indirect effects (Preacher, Rucker, & Hayes, 2007: 198–199).

RESULTS

Descriptive results (displayed in Table 1) indicate a positive correlation between relative duration of a surgery and the number of adverse events ($r = .24$, $p < .01$). However, this relationship becomes nonsignificant when complexity is taken into account. In addition, we found an unexpected negative correlation between ATL and team helping ($r = -.22$, $p < .05$). Still, as noted below (see Table 4, model 10), this relationship becomes positive and significant when we control for ASA. The analysis of the baseline (i.e., preintervention) data indicates no significant difference in either adverse events ($\chi^2_8 = 1.00$, $p = .99$) or relative duration ($F[8, 89] = 0.2$, $p = .99$) among the teams from the different wards regardless of condition.

Even though not all teams in the intervention condition consistently participated in "post-op" debriefings, an analysis of the postintervention data indicates a main effect of the intervention on relative duration (mean relative duration, control = 1.12, mean relative duration, intervention = 0.89, $t_{222} = 2.76$, $p < .01$), albeit not on adverse events ($\chi^2_{240} = 0.02$, n.s.).

The control variables were not significantly related to the two dependent variables, except for a positive and significant relationship between team size and surgical duration, as can be seen in the first model (model 1) presented in both Tables 2 and 3. Model 2, Table 2, and model 2, Table 3, examine the main effects of ATL (Hypotheses 1a, 1b). As can be seen in Table 2's model 2, ATL had a significant negative effect ($b = -0.02$, $p < .05$) on surgical duration after the control variables are taken into consideration. This model differed significantly from a model including only the control

TABLE 1
Descriptive Statistics[a]

Variable	*n*	Mean	s.d.	1	2	3	4	5	6	7	8	9
1. ASA	241	2.24	0.88									
2. Briefing	250	0.37	0.48	.28***								
3. Team size	244	5.00	1.39	.01	−.03							
4. Centered number of surgeries in which at least two members participated in together	250	0	6.15	−.07	−.10	.34***						
5. Complexity	227	1.88	0.82	.29***	.15*	.25***	.11					
6. ATL	244	5.63	6.55	.052	.22***	.09	.47***	−.14*				
7. Team workload sharing	102	5.66	0.85	−.14	.21*	−.01	−.15	−.09	.20*			
8. Team helping	119	5.54	1.27	−.15	.18*	−.00	−.11	−.03	−.22*	.59***		
9. Relative duration of surgery	248	1.00	0.67	.12	.01	.23***	.08	.58***	−.10	−.25*	−.13	
10. Number of adverse events	250	0.60	0.82	.19**	.05	.17**	.03	.33***	−.06	.04	.06	.24***

[a] "ASA" is the American Society of Anesthesiologists index. "ATL" is action team learning.
* $p < .05$
** $p < .01$
*** $p < .001$

TABLE 2
Results of RCM Analysis for Relative Duration as Dependent Variable[a]

Variables	Model 1: Control		Model 2: Direct Effect		Mediation: Model 3a		Mediation: Model 3b		Moderated Mediation: Model 4a		Moderated Mediation: Model 4b	
n	205		205		73		63		73		63	
Effect												
Intercept	0.27	(0.23)	0.42	(0.22)	1.83**	(0.67)	1.35	(0.80)	−0.67	(0.93)	−1.36	(1.29)
ASA	0.15**	(0.05)	0.14**	(0.05)	0.21*	(0.10)	0.21	(0.12)	0.15	(0.09)	0.08	(0.09)
Briefing	0.13	(0.10)	0.11	(0.09)	−0.01	(0.18)	0.001	(0.23)	0.12	(0.17)	0.32	(0.19)
Team size	0.08*	(0.03)	0.07*	(0.03)	0.06	(0.06)	0.03	(0.07)	0.01	(0.05)	−0.01	(0.05)
Centered number of surgeries in which at least two members participated	0.01	(0.01)	0.02*	(0.01)	0.03	(0.02)	0.04*	(0.02)	0.01	(0.01)	0.01	(0.02)
ATL			−0.02*	(0.01)	−0.02	(0.02)	−0.02	(0.02)	0.003	(0.02)	0.04	(0.03)
Team workload sharing					−0.22*	(0.11)			0.10	(0.15)		
Team helping							−0.11	(0.10)			0.22	(0.19)
High complexity (vs. low)									3.36**	(1.15)	3.18*	(1.30)
Medium complexity (vs. low)									3.76**	(1.22)	4.09**	(1.31)
Team workload sharing × complexity (medium vs. low)									−0.44*	(0.20)		
Team workload sharing × complexity (high vs. low)									−0.41*	(0.21)		
Team helping × complexity (medium vs. low)											−0.39	(0.20)
Team helping complexity (high vs. low)											−0.42*	(0.21)
ATL × high complexity									−0.03	(0.02)	−0.06	(0.03)
ATL × medium complexity									0.02	(0.03)	0.01	(0.03)
Random variance, ward	0.02	(0.04)	0		0		0					
Random variance, head surgeon	0.07*	(0.04)	0.07*	(0.03)	0.13*	(0.07)	0.19*	(0.09)	0.16	(0.12)	0.26	(0.18)
Random variance, complexity by head surgeon									0.07	(0.05)	0.06	(0.04)
−2 log-likelihood	356.5		352.3		145.5		134.1		99.9		83.8	
Δ−2 log-likelihood[b]			4.2*						45.6***			

[a] Values are unstandardized coefficients with standard errors in parentheses. "ASA" is the American Society of Anesthesiologists index. "ATL" is action team learning.
[b] Calculation based on a main-effect model with the same sample size and same random variance term as the moderation model.
* $p < .05$
** $p < .01$
*** $p < .001$

variables (Δ−2 log-likelihood = 4.2, $p < .05$). The random variance between head surgeons was significant in this model. In contrast, ATL had no significant direct effect on adverse events.

Hypotheses 2a and 2b posit that team workload sharing partially mediates the relationship between ATL and both performance-related team outcomes. As workload sharing was measured by aggregating team members' responses to a questionnaire immediately following surgery, our sample size for the analysis for these hypotheses dropped substantially. Only 73 teams (43 in the intervention condition and 30 in the control condition, with an average of approximately 1.5 people answering in each team) met the inclusion criteria (no missing data with respect to the dependent and independent variables). This response rate lies within the boundaries found acceptable by Baruch (1999) when studying physicians. To ensure the absence of sample bias, we compared these 73 teams with those excluded from the analysis with respect to all relevant variables. We found no significant difference between the two sets of teams with respect to either relative duration ($mean_{73}$ = 1.13, $mean_{175}$ = 0.94, t = 1.84, n.s.), occurrence of adverse events ($mean_{73}$ = 0.63, $mean_{175}$ = 0.59, χ^2 = 0.14, n.s.), or task complexity ($mean_{73}$ = 1.85, $mean_{153}$ = 1.68, t = 1.97, n.s.). In addition, we conducted a sensitiv-

TABLE 3
Analysis Examining the Main Effect of Action Team Learning and Moderation of Complexity on Number of Adverse Events[a]

Variables	Model 1		Model 2		Model 3	
Effect						
Intercept	−1.62**	0.54	−1.57**	0.54	−1.81*	0.54
ASA	0.15	0.12	0.15	0.12	0.12	0.12
Briefing	−0.01	0.23	0.02	0.23	0.01	0.21
Size	0.15	0.08	0.15	0.08	0.12	0.08
Centered number of surgeries in which at least two members participated	−0.003	0.01	0.002	0.02	−0.02	0.02
ATL			−0.01	0.02	−0.05	0.03
High complexity (vs. low)					0.69*	0.34
Medium complexity (vs. low)					0.29	0.34
ATL × high complexity					0.06	0.04
ATL × medium complexity					0.08*	0.04
Random variance, ward	0.07	0.08	0.07	0.08	0	
Random variance, head surgeon	0		0		0	
−2 log-pseudo-likelihood[b]	701		701		715.02	

[a] $n = 199$. "ASA" is the American Society of Anesthesiologists index. "ATL" is action team learning.
[b] The log-pseudo-likelihood in a GLMM is the log-likelihood of a linearized model. One should not compare these values across different statistical models, even if the models are nested with respect to fixed and/or G-side random effects. It is possible that between two nested models the larger model has a smaller pseudo-likelihood.
* $p < .05$
** $p < .01$

ity analysis to examine whether within-team response rate affected our results. We calculated the proportion of team members responding to the questionnaire in each team and added this variable to our models as an additional moderator. The interaction of ATL or ATL by complexity with this variable was nonsignificant in all of our models.

To test for the mediating effects of workload sharing on the link between ATL and both performance-related team outcomes, we first assessed the main effect of ATL on workload sharing (path a of the mediation as depicted in Figure 1). As can be seen in model 1 of Table 4, this relationship was positive and significant ($b = 0.06$, $p < .01$). As can be seen in model 3a of Table 2, when both ATL and workload sharing are included as predictors of relative surgical duration, only workload sharing is significant ($b = -0.22$, $p < .05$). We assessed the significance of this apparent mediation by examining the significance and relative size of the indirect effect of ATL on duration via workload sharing ($a \times b = -0.01$). We applied a bootstrap procedure with the low/high 0.05 confidence interval (−0.03,−0.0002), indicating a significant indirect effect. We found that the indirect effect of ATL on duration via workload sharing accounted for 35 percent of the total effect of ATL on relative duration. Thus, Hypothesis 2b was supported. In contrast, we found no evidence of mediation by workload sharing with respect to the number of adverse events.

Hypotheses 2c and 2d posit that team helping partially mediates the relationship between ATL and both performance-related outcomes. Again, our sample size dropped substantially when we were conducting the analysis for these hypotheses. Only 63 teams (39 in the intervention condition and 24 in the control condition, with an average of approximately 1.5 people answering in each team) met the inclusion criteria noted earlier. Still, this response rate lies within the range of acceptability for physician-based research noted by Baruch (1999).

To test for the mediating effects of team helping on the link between ATL and the performance-related outcomes, we first assessed the main effect of ATL on helping (path a of the mediation as depicted in Figure 1). As can be seen in model 2 of Table 4, this relationship was positive and significant ($b = 0.06$, $p < .05$). However, helping was not found to mediate the relationship between ATL and either of the performance measures.

Hypotheses 3a and 3c posit that task complexity moderates the relationship between team workload sharing and the performance-related outcomes. To test these hypotheses, we expanded model 3a to include two dummy variables—high versus low

TABLE 4
Results of Analysis to Examine the Main Effect of ATL on Team Processes[a]

Dependent Variables	Model 1: Team Workload Sharing		Model 2: Team Helping	
Effect				
Intercept	5.14***	0.40	6.42***	0.58
ASA	−0.13	0.1	−0.31*	0.14
Briefing	0.28	0.18	0.29	0.29
Size	0.07	0.06	−0.09	0.09
Centered number of surgeries in which at least two members participated	−0.06**	0.02	−0.06*	0.03
ATL	0.06**	0.02	0.06*	0.03
Team workload sharing				
Random variance, ward	0		0	
Random variance, head surgeon	0		0.33	0.26
−2 log-likelihood	158.3		160.8	

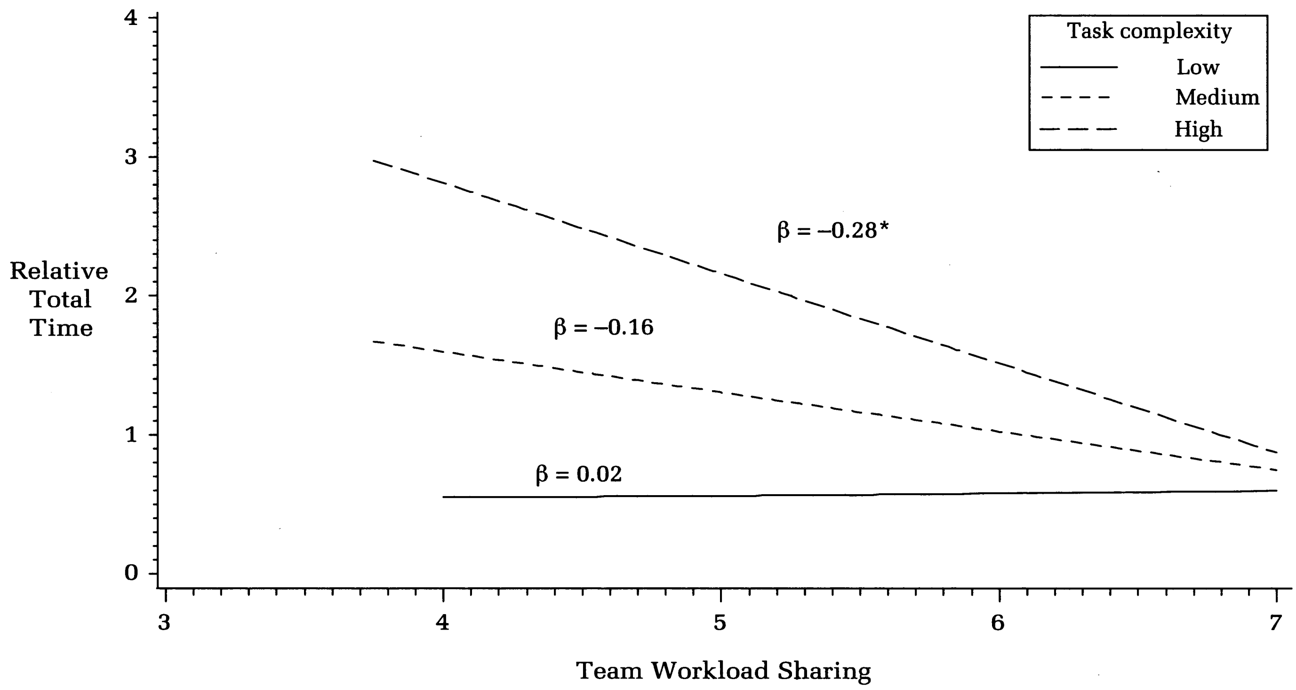
[a] $n = 73$ for model 1; $n = 63$ for model 2. "ASA" is the American Society of Anesthesiologists index. "ATL" is action team learning.
* $p < .05$
** $p < .01$
*** $p < .001$

complexity and medium versus low complexity—as well as the interactions between workload sharing and these dummy variables, and between ATL and these dummy variables. Hypothesis 3a (regarding the complexity-moderated effect of workload sharing on adverse events) was not supported, in that neither of the two workload sharing–complexity interaction terms were statistically significant (results are available from the authors). However, as shown in model 4a of Table 2, Hypothesis 3c (regarding the complexity-moderated effect of workload sharing on duration) *was* supported. The interactions of workload sharing with both low versus medium complexity and low versus high complexity were significant ($b = -0.44$, $p < .05$; $b = -0.41$, $p < .05$). This indicates that the interaction between workload sharing and complexity mediates the effect of ATL on duration. The change in the −2 log-likelihood ($\Delta = 45.6$, $p < .001$) between models 3a and 4a indicates that the inclusion of the complexity–workload sharing interaction significantly contributes to the model's explanatory potential. Moreover, as indicated by the simple slope analysis (Figure 2), the beneficial effects of workload sharing with respect to surgical duration are, as hypothesized, most robust at high levels of task complexity ($b = -0.28$, $p < .05$).

Hypotheses 3b and 3d posit that complexity moderates the relationship between helping and both relative duration and the number of adverse events. To test these hypotheses and the ones following, we examined a model including the direct effect of ATL, helping, task complexity, the interaction between team helping and task complexity, and the interaction between ATL and task complexity. Hypothesis 3b (regarding the complexity-moderated effect of helping on adverse events) was not supported in that neither of the two helping-complexity interaction terms was statistically significant (results are available from the authors). However, as shown in model 4b of Table 2, Hypothesis 3d, regarding relative duration, *was* supported. The interaction between helping and low versus high complexity was significant ($b = -0.42$, $p < .05$), indicating that the interaction between helping and complexity mediates the effect of ATL on relative duration. The change in the −2 log-likelihood ($\Delta = 50.3$, $p < .001$) between models 3b and 4b of Table 2 indicates that the inclusion of the complexity-helping interaction significantly contributes to the model's explanatory potential. As is evident from the simple slope analysis (Figure 3), the beneficial effects of helping on relative duration are, as hypothesized, more robust at high levels of complexity ($b = -0.22$, $p < .05$). Moreover, at low levels of complexity team helping had the opposite effect, with more helping related to *longer* relative duration ($b = 0.35$, $p < .001$)

Hypothesis 4a posits that team task complexity moderates the direct association between ATL and the number of adverse events. As we did not find any evidence of mediation with respect to adverse events, we tested the moderated-direct effect of ATL in a separate model that did not include any mediators. Model 3 of Table 3 shows a significant interaction of ATL with medium versus low levels of task complexity ($b = 0.08$, $p < .05$). Moreover, as seen in Figure 4, the nature of the interaction is consistent with the hypothesis. In low-complexity surgeries, the greater the level of ATL, the fewer adverse events occur. Moreover, under such low-complexity conditions, the effect associated with varying levels of ATL is substantial. For example, under low-complexity conditions, a one standard deviation increase in ATL is associated with a 5 percent reduction in the rate of adverse events (though this decrease is only marginally significant; $p < .1$). In contrast, the level of ATL was not associated with the number of adverse events in

FIGURE 2
Interaction between Workload Sharing and Complexity with Relative Duration as Dependent Variable



* $p < .05$

either medium- or high-complexity surgeries (b = 0.02, n.s. for both).

No support was found for Hypothesis 4b, which posits that team task complexity moderates the relationship between ATL and relative duration.

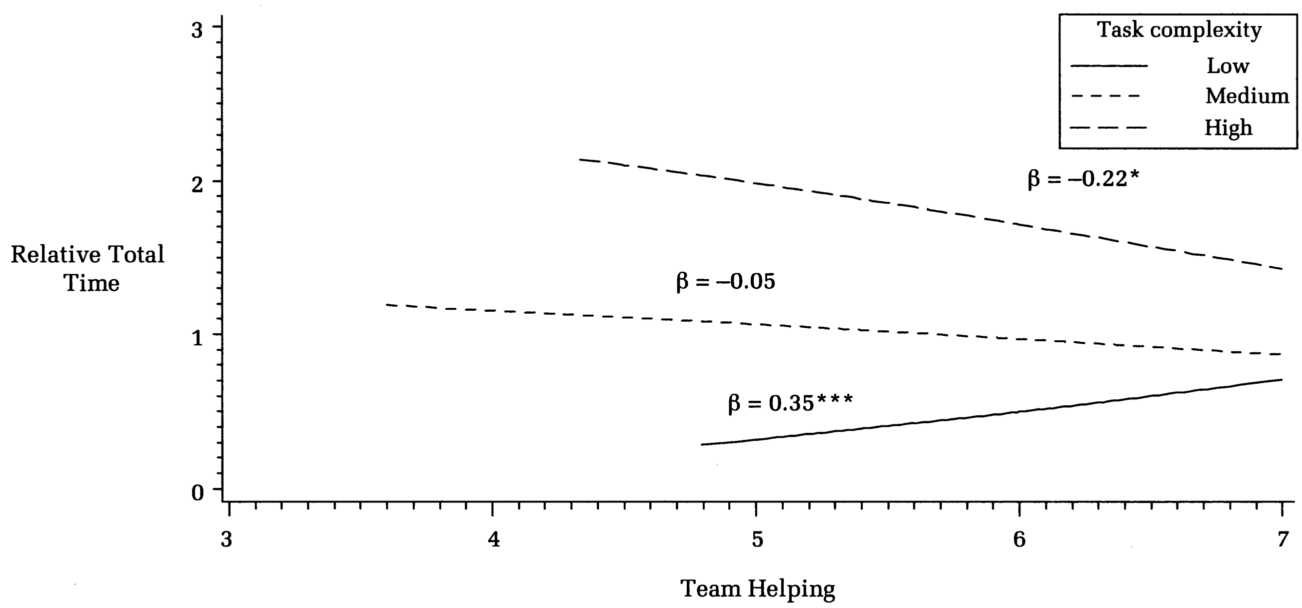
DISCUSSION

Drawing from and extending prior research on team reflexivity (e.g., Gurtner et al., 2007; Schippers et al., 2007; Smith-Jentsch et al., 2008; West, 1996, 2000), we argued that the regular and continuous guided team reflexivity experiences of an action team's members may serve as a viable substitute for naturally occurring, long-term learning processes in stable teams. More specifically, we proposed that aside from any *direct* benefits of such action team learning on team performance, such learning would have important *indirect* effects on team performance via its impact on coordination-based team processes, so that members of teams with higher levels of ATL would be better able to anticipate other role holders' behavior and more rapidly and effectively adapt their responses to each others' actions and to the dynamic circumstances that drive them. Moreover, we proposed that the level of team task complexity would condition these direct and indirect effects.

Overall, our findings generally support our contention that the compositional instability and short life span inherent to action teams need not impede team learning. Focusing first on the direct, unmoderated effects of ATL, we found that although a higher level of ATL is not associated with any reduction in adverse events, it is indeed associated with shorter surgical duration (i.e., greater efficiency). Moreover, these performance-related benefits of ATL are independent of and go beyond any effect offered by members' participation in a premission action team briefing session (of the type analyzed by Gurtner et al. [2007]).

Beyond showing merely that temporally unstable teams such as action teams can learn, we also sought to specify *how* and *when* such team learning yields performance-related benefits, and by doing so, contribute to the development of a midlevel theory of learning in such teams. In terms of "how," we posited two coordination-based constructs as likely to explain the impact of such learning on team performance. Drawing on Hackman's (1990,

FIGURE 3
Interaction between Team Helping and Complexity with Relative Duration as Dependent Variable



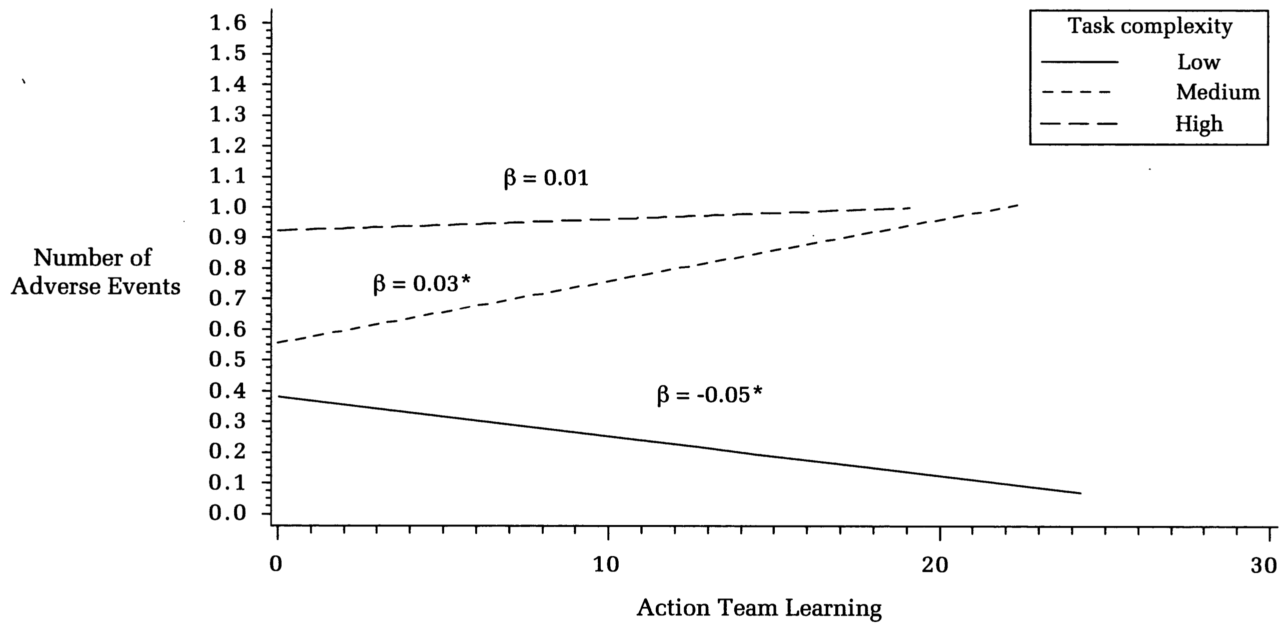
* $p < .05$

1993) general notion that team effectiveness is largely a function of teammates' cooperation and coordination, we proposed that workload sharing and helping behaviors mediate the effect of ATL on team performance. We based this prediction on the idea that with higher levels of ATL, a team's members develop (1) a better cross-understanding of what others expect from them in their team role, as well as what they may expect from those filling other team roles (Huber & Lewis, 2010), and (2) a better implicit sense of what may go wrong during team operations, how other role holders will respond to such situations, and what actions they should take to maximize the collective response of the team. As such, we proposed that ATL facilitates implicit coordination among team members, manifested in enhanced workload sharing and helping (Rico et al., 2008).

We found partial support for these mediation hypotheses. Workload sharing mediated the association between ATL and surgical duration, though helping did not. Neither variable mediated the association between ATL and adverse events. We believe that this latter finding may be a result of type II error, given the low base rate of adverse events and the reduced number of observations for our mediation analysis. With a larger sample size, implicit coordination processes may still be found to mediate the ATL-adverse events relationship.

The limited findings with regard to the mediating role of workload sharing and helping vis-à-vis surgical duration are not entirely surprising, given that we hypothesized that these relationships are contingent upon the complexity of team tasks. Specifically, we theorized that the implicit coordination elicited by ATL may have a stronger impact on performance under conditions of greater team task complexity. We indeed found support for a moderating effect of complexity on the relationship between the two implicit coordination variables and surgical duration. The results of our simple slopes analyses suggest that although ATL has significant beneficial effects on both implicit coordination processes, these coordination processes are themselves associated with shorter surgical duration only when team tasks are more complex. When team tasks are less complex, these processes appear to have a diminished effect on duration. This makes sense, because less complex surgeries tend to be shorter in duration, thus limiting the possibility that any beneficial team processes will be able to shorten them further. In addition, in keeping with classic contingency theory (Thompson, 1967), implicit coordination may simply be less

FIGURE 4
Interaction between Action Team Learning and Complexity with Number of Adverse Events as Dependent Variable



* $p < .05$

relevant when means-ends relations are clearly understood, procedures are clear and simple, and interdependencies among team members are less intensive. Indeed, our results—that is, that workload sharing mediated the ATL-duration association, but helping did not—suggest that helping in less complex surgeries may in fact complicate standardized work processes, thereby offsetting any temporal benefits.

While we expected that task complexity would amplify the association between team coordination processes and team performance, we also theorized that task complexity would *attenuate* the *direct* effects of ATL on both performance outcomes. We based this prediction on the assumption that the unmediated effect of ATL on performance at least partially reflects domain-specific knowledge gleaned in prior team reflexivity experiences and on the notion proposed by O'Leary et al. (2011) that such domain knowledge is only generalizable to relatively analogous situations. And consistently with this notion, we found the direct, beneficial effect of ATL on adverse events to be stronger in low-complexity surgeries and weaker in surgeries of greater complexity, where prior domain knowledge can be assumed to be less relevant. These findings are important for two reasons. First, although others have suggested the potential impact of reflexivity on the enhancement of teams' content or domain knowledge, we are the first to attempt to disentangle these process and content/domain effects. Although we did not directly assess the latter, by demonstrating that two primary process factors (i.e., helping and workload sharing) only *partially* mediate the ATL-performance relationship, our findings are consistent with and suggestive of such a content/domain argument. Second, by demonstrating that team task complexity moderates the direct effect of ATL on adverse events, we provide some of the first (albeit, indirect) empirical support for O'Leary et al.'s (2011) argument that the transferability of domain knowledge from one team to another is complexity-contingent.

Interestingly, we found no evidence of a complexity-moderated *direct* effect with regard to relative duration. Both the moderated and unmoderated *direct* effects of ATL on surgical duration are nonsignificant when coordination-based mediators are taken into account. This suggests that the effect of ATL on efficiency-related outcomes may be attributable chiefly to enhancements in team processes, rather than to the domain knowledge of team members.

Overall, our moderation analyses show that, depending upon the level of task complexity, ATL can have important benefits for action teams and those they serve. More specifically, our findings suggest that greater ATL can reduce the number of adverse events in low-complexity team tasks, albeit at the cost of increased mission duration (owing to the offsetting effects of greater team helping on time savings in low-complexity team tasks). Further, our findings suggest that in more complex tasks, greater ATL-induced workload sharing and helping can improve efficiency (i.e., reduce relative duration) without exacting a price in the number of adverse events.

Taken together, our study and its findings offer a number of important contributions to team learning theory. First, our study suggests an alternative approach to conceptualizing learning in temporally unstable teams such as action teams. According to this approach, team learning in such teams is viewed as a team-level property capturing the degree to which a team's members accumulate and bring to that team a body of *cross-team* reflexivity experience allowing members to more immediately and effectively anticipate and coordinate action, even for the relatively brief time that they may be working with each other. Such an approach is an important departure from the more conventional approaches to framing and studying team learning reviewed by Edmondson et al. (2008) and Salas et al. (2007), all of which, by assuming team temporal stability, offer limited applicability to action teams.

Building on such a conceptualization of team learning, our study offers an important theoretical contribution by suggesting that in teams lacking temporal stability, such as action teams, the collective reflection and experience sharing that naturally occur either implicitly or explicitly in long-term, stable teams can be deliberately engineered using guided reflection, thus enabling a kind of transitive learning to occur as members move from one team to another. Although theoretical treatments of multiteam memberships (e.g., O'Leary et al., 2011) have highlighted the potential importance of such transitive learning processes in teams, our study extends this notion in two important ways. First, by drawing from the literature on after-event reviews (Ellis & Davidi, 2005), our study suggests that guided reflexivity may serve as a mechanism by which individual team members can generate shared understandings of team experiences even under conditions of high temporal *in*stability. This is important because, to date, an assumption of multiteam membership models of team learning has been target team membership stability, with members integrating knowledge gleaned from membership in other teams on the basis of more conventional, long-term team learning processes. Second, our study extends the multiteam membership notion of transitive learning in teams by explicating the conditions that are likely to affect the efficacy of such guided reflexivity when a target team is characterized by a high degree of temporal instability, as in the case of action teams. These conditions are that such learning is role-based (to ensure generalizability from one team to the next), and consistent (to allow for the identification of the common patterns of member interaction across teams). This is important in that it highlights for researchers and practitioners alike that *the way in which reflexivity is structured* may have important implications for team members' ability not only to learn from their own experience, but also to synthesize those insights with those gleaned by others in completely different team contexts.

Aside from demonstrating that such transitive forms of learning may have important performance-related implications for action teams, a second theoretical contribution offered by our study is its specification of some of the coordination-based mechanisms underlying such effects. More specifically, in keeping with the input-process-outcome approach (Guzzo & Dickson, 1996; Mathieu et al., 2000) and Salas et al.'s (2007) conclusion that enhanced coordination processes underlie the effect of team training on performance, our findings identify two coordination-based team processes—workload sharing and helping—as key mechanisms linking members' cumulative reflective experience to more effective team performance, as evidenced by the tendency of teams with higher levels of ATL to complete their missions in less time. We view this as an important contribution in that, by shedding light on *how* ATL may be linked to team performance, our model begins to fill the void in midlevel theories of team learning noted by Edmondson et al. (2007) in their recent review.

Finally, our study further contributes to such midlevel theory on team learning by also shedding light on *when* ATL may be linked to team performance. More specifically, our findings demonstrate that team task complexity may serve as an important boundary condition, moderating both the direct and indirect effects of ATL on team performance. This is important for two reasons. First, as others have suggested (DeRue et al., 2012; Edmondson et al., 2007; Salas et al., 2007), research on team learning is still at a relatively early stage, with most studies still aimed at trying to identify the various factors directly affecting team learning as a process or outcome, and little theory specifying how individual, team, or contextual factors may condition these influences. Indeed, while DeRue et al.'s recent study (2012) provides important insights into how individual differences may moderate the effect of guided reflexivity on team leadership development, to the best of our knowledge, ours is the first study to generate and test theory regarding the moderating role of team task context on reflexivity's direct and indirect performance-related consequences. Second, while our findings suggest that team task complexity moderates the effects of ATL on performance both directly and indirectly (i.e., by moderating the impact of coordination-related team processes on performance), to the extent that these same coordination processes (i.e., workload sharing and helping) likely mediate the impact of more conventional and less deliberate forms of learning in more temporally stable teams, our findings suggest the need to give greater consideration to the conditioning role of team task complexity not only in models of action team learning, but in models of team learning overall.

Beyond these important theoretical contributions, this study represents one of the first field intervention studies in the area of team-based learning and so offers an important empirical contribution. The vast majority of studies examining team learning and related concepts (e.g., team transactional memory, after-event reviews) are ethnographic (Edmondson, 1999), cross-sectional (Sexton, Thomas, & Helmreich, 2000), or laboratory experiments, in which students or soldiers engage in some sort of simulation (e.g., Ellis, 2006; Gurtner et al., 2007; Smith-Jentsch et al., 2007) or perform a simple assembly task (Lewis et al., 2005). Moreover, as DeRue et al. (2012: 2) noted, studies of after-event reviews are typically of short duration (e.g., Gurtner et al., 2007), in most cases tracking the impact of one or two rounds of guided reflexivity over short periods of time (i.e., less than one hour to three weeks). Additionally, the current study offers an important empirical contribution with respect to its investigation of learning in action teams. Although Klein et al. (2006) examined action teams in a field setting, their focus was leadership and not team learning. Indeed, we are unaware of any study that, using a semiexperimental, prospective design with randomized assignment, has sought to document the longitudinal effects of a learning-based intervention on teams (no less, action teams) in an actual work setting. In this sense, our findings lend external validity to earlier team learning studies by demonstrating the impact (albeit context-contingent) of learning on team performance.

Finally, our findings have obvious relevance for managerial practice, particularly for those managing action teams in health care settings, in that the findings show that postaction debriefings may constitute an exemplary process for enhancing learning in action teams. The importance of such learning in hospital-based action teams cannot be overstated, given our findings that ATL is associated with significantly shorter surgical duration (particularly in more complex surgeries) and a significant decline in the occurrence of adverse events under conditions of low task complexity—precisely when, according to Regenbogen, Greenberg, Studdert, Lipsitz, Zinner, and Gawande (2007), such events tend to occur.

Limitations and Future Research

Our study has a number of theoretical and methodological limitations. First, it may be argued that unweighted cumulative guided reflexivity experience (rather than cumulative experience as weighted by the regularity of that experience) serves as the best indicator of ATL. With this in mind, we reran the analyses with team members' cumulative reflexivity experience as the only independent variable and members' total surgical experience as a control variable. Our results were unchanged with regard to relative duration, but they were weaker with regard to the number of adverse events (the interaction of ATL with task complexity was only significant at the .10 level). This finding supports our contention that the regularity of reflective experiences plays a key role in ATL and is also consistent with Hertwig and Erev's (2009) theory of rare events, which suggests that people tend to underestimate the likelihood of rare adverse events, particularly when reflexivity is discontinuous. To the extent that action team members take part in guided reflexivity on a more regular basis, the theory of rare events suggests that they will be better able to recognize links between a new rare event and previously experienced situations, to understand the factors contributing to such events, and to more effectively develop team-based strategies for preventing such events or responding to them when they occur.

Second, while we focused on coordination behaviors (workload sharing and helping) as the key mechanisms mediating ATL's impact on team performance, our findings lent only partial support to our hypotheses in this regard. Empirical limitations (such as limited statistical power in the models specifying mediation) may underlie the absence of more complete support. However, other team processes, such as team members' vigilance or attention to detail (Miron, Erez, & Naveh, 2004), may also help mediate this relationship. Accordingly, we encourage researchers to investigate the possible role of these and other processes in mediating the effects of ATL on performance-related team outcomes. In addition, it may be argued that while we conceptualized workload sharing and helping as coordination mechanisms, they may in fact be reflective of team commitment or engagement, with these constructs rather than coordination serving as the primary mediator of ATL's impact on performance. Thus we also encourage research aimed at directly testing the possible mediating role of team commitment and engagement as mediators of the ATL-performance relationship.

A third limitation stems from our single-item assessment of team task complexity. It is possible that in responding to this item, participants may have focused more on the degree of interdependence required by the task at hand rather than the uncertainty inherent in it. If this is so, consistently with relational cohesion theory (Lawler, Thye, & Yoon, 2008), our results may reflect the influence of team interdependence on implicit cooperation more than the interaction between task complexity and ATL. To rule out the possible confounding effects of interdependence, we ran a sensitivity analysis (Kim, 1984) replacing participants' assessments of the complexity of the surgery just completed with the anesthesiologists' ratings of generic surgical complexity noted earlier, and incorporating the two-item subscale for coordinative complexity reflecting interdependence (see the Appendix) as a control. The inclusion of interdependence as a control had no meaningful impact on the findings. While the magnitude of some of the parameter estimates shifted slightly, their relative magnitudes remained identical to those generated via the models discussed above, and there was no impact whatsoever on the statistical significance of the estimates. Nevertheless, given that the complexity concept is multidimensional, tapping not only uncertainty and interdependence but also the manner in which tasks are divided among team members, we encourage research aimed at exploring how each of these unique but correlated factors may moderate the impact of ATL on team performance outcomes.

Fourth, while we used adverse events and relative duration as indicators of quality and efficiency respectively, these measures may not always provide an accurate reflection of these two performance outcomes. For example, surgical teams may be willing to allow certain correctable adverse events to occur if they facilitate other, more critical steps in a patient's care. Additionally, simply counting the number of adverse events may provide a biased impression of quality, as some adverse events may be more serious and more difficult to address than others. One very serious adverse event may be less reflective of quality than a high number of minor, easily addressed adverse events. Accordingly, we encourage researchers to test our hypotheses across a broader range of team outcomes.

Fifth, we would be remiss were we to leave untested the alternative explanation that it is not team ATL that affects team performance, but rather, the expertise or leadership qualities of a team's leader (i.e., here, the head surgeon). To test this possibility, for each head surgeon in the three intervention wards, we calculated the percent of surgeries led by that surgeon in which a reflexivity session was conducted. We then examined the correlation between this variable and the dependent variables of our study (i.e., relative duration and number of adverse events) for all teams in the intervention wards. We found no significant correlation between a head surgeon's tendency to lead reflexivity sessions and the relative duration of surgeries led by that surgeon ($r = -.14$, n.s.), and a *positive* correlation between the head surgeon's tendency to lead reflexivity sessions and the number of adverse events ($r = .18$, $p < .05$). This suggests that leadership qualities of a head surgeon, as operationalized by a tendency to lead reflexivity sessions, are not in themselves related to a reduction in surgery time or the number of adverse events.

Aside from the research challenges noted above, we encourage research aimed at assessing the degree to which our conceptualization of ATL can be generalized to action teams operating in other contexts. Similarly, research is needed to assess the context specificity of our findings regarding how and when ATL is linked to team performance (Bamberger, 2008).

Finally, while our findings highlight the potential value of transitive team learning for action teams, this does not mean that such learning has applicability only to action teams. Indeed, as O'Leary et al. (2011) suggested, such cross-team learning is likely to take on increased salience not merely as work becomes increasingly team-based, but also as it becomes increasingly structured around temporally unstable teams and multiple team memberships. Although our study provides a basic template for understanding how and when team learning might be enhanced for *teams* whose members have multiple and shifting affiliations, there is still far more for *us* as researchers to learn.

REFERENCES

Adler, P., Goldoftas, B., & Levine, D. I. 1999. Flexibility versus efficiency?: A case study of model changeovers in the Toyota production system. ***Organization Science,*** 10: 43–68.

Altman, D. G. 1991. ***Practical statistics for medical research.*** London: Chapman and Hall.

Anderson, S. E., & Williams, L. J. 1996. Interpersonal, job, and individual factors related to helping processes at work. ***Journal of Applied Psychology,*** 81: 282–296.

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

Argyris, C. & Schön, D. A. 1978. ***Organizational learning: A theory of action perspective.*** Reading, MA: Addison-Wesley.

Bamberger, P. A. 2009. Employee help-seeking: Antecedents, consequences and new insights for future research. ***Research in Personnel and Human Resources Management,*** 28: 49–98.

Bamberger, P. 2008. Beyond contextualization: Using context theories to narrow the micro-macro gap in management research. ***Academy of Management Journal,*** 51: 839–846.

Baruch, Y. 1999. Response rate in academic studies: A comparative analysis. ***Human Relations,*** 52: 421–438.

Bergami, M., & Bagozzi, R. P. 2000. Self-categorization, affective commitment and group self-esteem as distinct aspects of social identity in the organization. ***British Journal of Social Psychology,*** 39: 555–577.

Blickensderfer, E., Cannon-Bowers, J. A., & Salas, E. 1997. Theoretical bases for team selfcorrection: Fostering shared mental models. In M. M. Beyerlein, D. A. Jackson, & S. T. Beyerlein (Eds.), ***Advances in interdisciplinary studies of work teams:*** 249–279. Greenwich, CT: JAI.

Bryk, A., & Raudenbush, S. 1992. ***Hierarchical linear models: Applications and data analysis methods.*** Newbury Park, CA.: Sage.

Carter, S., & West, M. A. 1998. Reflexivity, effectiveness, and mental health in BBC-TV production teams. ***Small Group Research,*** 29: 583–601.

Chudoba, K. M., & Watson-Manheim, M. B. 2007. Exploring the virtual work environment: A process perspective. In A. J. Salazar & S. Sawyer (Eds.), ***Handbook of information technology in organizations and electronic markets:*** 57–76. Singapore: World Scientific.

Davenport, D. L., Bowe, E. A., Henderson, W. G., Khuri, S. F., & Mentzer, R. M. 2006. National Surgical Quality Improvement Program (NSQIP) risk factors can be used to validate American Society of Anesthesiologists physical status classification (ASA PS) levels. ***Annals of Surgery,*** 243: 636–641.

DeRue, D. S., Nahrgang, J. D., Hollenbeck, J. R., & Workman, K. 2012. A quasi-experimental study of after-event reviews and leadership development. ***Journal of Applied Psychology,*** 97: 997–1015.

Edmondson, A. 1999. Psychological safety and learning behavior in work teams. ***Administrative Science Quarterly,*** 44: 350–383.

Edmondson, A., Dillon, J. R., & Roloff, K. S. 2008. Three perspectives on team learning: Outcome improvement, task mastery, and group process. In A. P. Brief & J. P. Walsh (Eds.), ***Academy of Management annals,*** vol. 1: 269–314. New York: Erlbaum.

Edwards, J. R., & Lambert, L. S. 2007. Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. ***Psychological Methods,*** 12: 1–22.

Ellis, A. P. J. 2006. System breakdown: The role of mental models and transactive memory in the relationship between acute stress and team performance. ***Academy of Management Journal,*** 49: 576–589.

Ellis, A. P. J., Bell, B. S., Ployhart, R. E., Hollenbeck, J. R., & Ilgen, D. R. 2005. An evaluation of generic teamwork skills training with action teams: Effects on cognitive and skill-based outcomes. ***Personnel Psychology,*** 58: 641–672.

Ellis, S., & Davidi, I. 2005. After event reviews: Drawing lessons from failed and successful experience. ***Journal of Applied Psychology,*** 90: 857–871.

Erev, I., & Haruvy, E. 2010. Two-stage lotteries and the value of unresolved uncertainty. ***Marketing Letters,*** 21: 149–162.

Erez, A., LePine, J. A., & Elms, H. 2002. Effects of rotated leadership and peer evaluations on the functioning and effectiveness of self-managed teams: A quasi-experiment. ***Personnel Psychology,*** 55: 929–948.

Fiore, S. M., Salas, E., & Cannon-Bowers, J. A. 2001. Group dynamics and shared mental model development. In M. London (Ed.), ***How people evaluate others in organizations:*** 309–336. Mahwah, NJ: Erlbaum.

Fleiss, J. L., Levin, B., & Paik, M. C. 2004. The measurement of interrater agreement. In ***Statistical methods for rates and proportions*** (3rd ed.): 598–676. Hoboken, NJ: Wiley.

Florin, J., Lubatkin, M., & Schulze, W. 2003. A social capital model of high-growth ventures. ***Academy of Management Journal,*** 46: 374–384.

Gawande, A. A., Thomas, E. J., Zinner, M. J., & Brennan, T. A. 1999. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. ***Surgery,*** 126: 66–75.

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

Goldstein, H. 1987. ***Multilevel models in educational and social research.*** London: Griffin.

Gurtner, A., Tschan, F., Semmer, N. K., & Nägele, C. 2007. Getting groups to develop good strategies: Effects of reflexivity interventions on team process, team performance and shared mental models. ***Organizational Behavior and Human Decision Processes,*** 102: 127–142.

Guzzo, R. A., & Dickson, M. W. 1996. Teams in organizations: Recent research on performance and effectiveness. In L. W. Porter & M. R. Rosenzweig (Eds.), ***Annual review of psychology,*** vol. 47: 307–338. Palo Alto, CA: Annual Reviews.

Haas, M. R. 2006. Knowledge gathering, team capabilities, and project performance in challenging work environments. ***Management Science,*** 52: 1170–1184.

Hackman, R. J. 1990. ***Groups that work (and those that don't): Creating conditions for effective teamwork.*** San Francisco: Jossey-Bass.

Hackman, R. J. 1993. Teams, leaders, and organizations: New directions for crew-oriented flight training. In E. L. Weiner, B. G. Kanki, & R. L. Helmreich (Eds.), ***Cockpit resource management.*** San Diego: Academic.

Hazelhurst, B., Mcmullen, C., & Gorman, P. 2004. Getting the right tools for the job: Distributed planning in cardiac surgery. ***IEEE Transactions on Systems, Man, and Cybernetics. Part A: Systems and Humans,*** 34: 708–717.

Hertwig, R., & Erev, I. 2009. The description–experience gap in risky choice. ***Trends in Cognitive Sciences,*** 13: 517–523.

Hollenbeck, J. R., Beersma, B., & Schouten, M. E. 2012. Beyond team types and taxonomies: A dimensional scaling conceptualization for team description. ***Academy of Management Review,*** 37: 82–106.

Horwitz, S. K., & Horwitz, I. B. 2007. The effects of team diversity on team outcomes: A meta-analytic review of team demography. ***Journal of Management,*** 33: 987–1015.

Huber, G. P., & Lewis, K. 2010. Cross understanding: Implications for group cognition and performance. ***Academy of Management Review,*** 35: 6–26.

Huey, B. M., & Wickens, C. D. (Eds.). 1993. ***Workload transition: Implications for individual and team performance.*** Washington, DC: National Academy Press.

Joshi, A., Pandey, N., & Han, G. 2009. Bracketing team boundary spanning: An examination of task-based, team level, and contextual antecedents. ***Journal of Organizational Behavior,*** 30: 731–759.

Kandel, E. R. 2007. ***In search of memory: The emergence of a new science of mind.*** New York: Norton.

Kim, J. 1984. An approach to sensitivity analysis in sociological research. ***American Sociological Review,*** 49: 272–282.

Klein, K. J., Ziegert, J. C., Knight, A. P., & Xiao, Y. 2006. Dynamic delegation: Shared, hierarchical, and deindividualized leadership in extreme action teams. ***Administrative Science Quarterly,*** 51: 590–621.

Kostopoulou, O., & Delaney, B. 2007. Confidential reporting of patient safety events in primary care: Results from a multilevel classification of cognitive and system factors. ***Quality and Safety in Health Care,*** 16: 95–100.

Lawler, E. J., Thye, S. R., & Yoon, J. 2008. Social exchange and micro social order. ***American Sociological Review,*** 73: 519–542.

Lawrence, P. R., & Lorsch, J. W. 1967. ***Organization and environment: Managing differentiation and integration.*** Homewood, IL: Irwin.

Lewis, K., Lange, D., & Gillis, L. 2005. Transactive memory systems, learning, and learning transfer. ***Organization Science,*** 16: 581–598.

Marks, M. A., Zaccaro, S. J., & Mathieu, J. E. 2000. Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. ***Journal of Applied Psychology,*** 85: 971–986.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. 2000. The influence of shared mental models on team process and performance. ***Journal of Applied Psychology,*** 85: 273–283.

Miron, E., Erez, M., & Naveh, E. 2004. Do personal characteristics and cultural values that promote innovation, quality, and efficiency compete or complement each other. ***Journal of Organizational Behavior,*** 25: 175–199.

Mohammed, S., & Dumville, B. C. 2001. Team mental models in a team knowledge framework. ***Journal of Organizational Behavior,*** 22: 89–106.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. 2005. When moderation is mediated and mediation is moderated. ***Journal of Personality and Social Psychology,*** 89: 852–863.

Ng, K. Y., & Van Dyne, L. 2005. Antecedents and performance consequences of helping behavior in work groups: A multi-level analysis. ***Group and Organization Management,*** 30: 514–540.

O'Leary, M. B., Mortensen, M., & Woolley, A. W. 2011. Multiple team membership: A theoretical model of its effects on productivity and learning for individuals and teams. ***Academy of Management Review,*** 36: 461–478.

Olivera, F., & Argote, L. 1999. Organizational learning and CORE processes in new product development. In L. Thompson, J. Levine, & D. Messick (Eds.), ***Shared cognition in organizations: The management of knowledge:*** 297–328. Hillsdale, NJ: Erlbaum.

Organ, D. W., Podsakoff, P. M., & MacKenzie, S. B. 2006. ***Organizational citizenship behavior: Its nature, antecedents, and consequences.*** Thousand Oaks, CA: Sage.

Podsakoff, N. P., Whiting, S. W., Podsakoff, P. M., & Blume, B. D. 2009. Individual- and organizational-level consequences of organizational citizenship behaviors: A meta-analysis. ***Journal of Applied Psychology,*** 94: 122–141.

Pope, C. 2002. Contingency in everyday surgical work. ***Sociology of Health and Illness,*** 24: 369–384.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. 2007. Assessing moderated mediation hypotheses: Theory, method, and prescriptions. ***Multivariate Behavioral Research,*** 42: 185–227.

Reagans, R., Argote, L., & Brooks, D. 2005. Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. ***Management Science,*** 51: 869–881.

Regenbogen, S. E., Greenberg, C. C., Studdert, D. M., Lipsitz, S. R., Zinner, M. J., & Gawande, A. A. 2007. Patterns of technical error among surgical malpractice claims: An analysis of strategies to prevent injury to surgical patients. ***Annals of Surgery,*** 246: 705–711.

Rico, R., Sanchez-Manzanares, M., Gil, F., & Gibson, C. 2008. Team implicit coordination processes: A team knowledge-based approach. ***Academy of Management Review,*** 33: 163–184.

Ron, N., Lipshitz, R., & Popper, M. 2006. How organizations learn: Post-flight reviews in an F-16 fighter squadron. ***Organization Studies,*** 27: 1069–1089.

Roth, E. M., Christian, C. K., Gustafson, M., Sheridan, T. B., Dwyer, K., Gandhi, T., Zinner, M. J., & Dierks, M. M. 2004. Using field observations as a tool for discovery: Analyzing cognitive and collaborative demands in the operating room. ***Cognition, Technology and Work,*** 6: 148–157.

Salas, E., Nichols, D., & Driskell, J. E. 2007. Testing three team training strategies in intact teams: A meta-analysis. ***Small Group Research,*** 38: 471–488.

Schippers, M. C., Den Hartog, D. N., & Koopman, P. L. 2007. Reflexivity in teams: A measure and correlates. ***Applied Psychology: An International Review,*** 56: 189–211.

Schippers, M. C., Den Hartog, D. N., Koopman, P. L., & Wienk, J. A. 2003. Diversity and team outcomes: The moderating effects of outcome interdependence and group longevity and the mediating effect of reflexivity. ***Journal of Organizational Behavior,*** 24: 779–802.

Schön, D. A. 1983. ***The reflective practitioner: How professionals think in action.*** New York: Basic Books.

Schön, D. A. 1987. ***Educating the reflective practitioner.*** San Francisco: Jossey-Bass.

Sexton, J. B., Thomas, E. J., & Helmreich, R. L. 2000. Error, stress, and teamwork in medicine and aviation: Cross sectional surveys. ***BMJ,*** 320: 745–749.

Shamir, B., & Kark, R. 2004. A single-item graphic scale for the measurement of organizational identification. ***Journal of Occupational and Organizational Psychology,*** 77: 115–123.

Smith, E. E., Nolen-Hoeksema, S., Fredrickson, B., & Loftus, G. 2003. ***Atkinson & Hilgard's introduction to psychology*** (14th ed.). Belmont, CA: Wadsworth/Thomson.

Smith-Jentsch, K. A., Campbell, G., Milanovich, D. M., & Reynolds, A. M. 2000. Measuring teamwork mental models to support training needs assessment, development, and evaluation: Two empirical studies. ***Journal of Organizational Behavior,*** 22: 179–194.

Smith-Jentsch, K. A., Cannon-Bowers, J. A., Tannenbaum, S. I., & Salas, E. 2008. Guided team self-correction: Impacts on team mental models, processes, and effectiveness. ***Small Group Research,*** 39: 303–327.

Spitzmuller, M., Van Dyne, L., & Ilies, R. 2008. Organizational citizenship behavior. A critical review and extension of its nomological network. In C. Cooper & J. Barling (Eds.), ***Handbook of organizational behavior:*** 106–123. London: Sage.

Staats, B. R., Gino, F., & Pisano, G. P. 2010. ***Varied experience, team familiarity, and learning: The mediating role of psychological safety.*** Working paper 10–016, Harvard Business School.

Stagl, K. C., Salas, E., & Day, D. V. 2008. Assessing team learning outcomes: Improving team learning and performance. In V. Sessa & M. London (Eds.), ***Work group learning: Understanding, improving, and assessing how groups learn in organizations:*** 367–390. Mahwah, NJ: Erlbaum.

Stasser, G., Stewart, D. D., & Wittenbaum, G. M. 1995. Expert roles and information exchange during discussion: The importance of knowing who knows what. ***Journal of Experimental Social Psychology,*** 31: 244–265.

Stewart, G. L., & Barrick, M. R. 2000. Team structure and performance: Assessing the mediating role of intra team process and the moderating role of task type. ***Academy of Management Journal,*** 43: 135–148.

Sundstrom, E., De Meuse, M. P., & Futrell, D. 1990. Work teams: Application and effectiveness. ***American Psychologist,*** 45: 120–133.

Thompson, J. D. 1967. ***Organizations in action: Social science bases of administration therapy.*** New York: McGraw-Hill.

Tushman, M. L. 1977. Special boundary roles in the innovation process. ***Administrative Science Quarterly,*** 22: 587–605.

Van der Vegt, G. S., Bunderson, S., & Kuipers, B. 2010. Why turnover matters in self-managing work teams: Learning, social integration, and task flexibility. ***Journal of Management,*** 36: 1168–1191.

van Ginkel, W., Tindale, R. S., & van Knippenberg, D. 2009. Team reflexivity, development of shared task representations, and the use of distributed information in group decision making. ***Group Dynamics: Theory, Research, and Practice,*** 13: 265–280.

Vashdi, D. R., Bamberger, P. A., Erez, M., & Weiss-Meilik, A. 2007. Briefing-debriefing: Using a reflexive organizational learning model from the military to enhance the performance of surgical teams. ***Human Resource Management,*** 46: 115–142.

Wanous, J. P., Reichers, A. E., & Hudy, M. J. 1997. Overall job satisfaction: How good are single-item measures? ***Journal of Applied Psychology,*** 82: 247–252.

Weingart, L. R. 1992. Impact of group goals, task component complexity, effort, and planning on group performance. ***Journal of Applied Psychology,*** 77: 682–693.

Weiss, H. M. 1990. Learning theory and industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), ***Handbook of industrial and organizational psychology,*** vol. 1: 171–221. Palo Alto, CA: Consulting Psychologists Press.

West, M. A. 1996. Reflexivity and work group effectiveness: A conceptual integration. In M. A. West (Ed.), ***The handbook of work group psychology:*** 555–579. Chichester, UK: Wiley.

West, M. A. 2000. Reflexivity, revolution, and innovation in work teams. In M. M. Beyerlein, D. A. Johnson, & S. T. Beyerlein (Eds.), ***Product development teams: Advances in interdisciplinary studies of work teams:*** 1–29. Stamford, CT: JAI.

West, M. A. 2002. Sparkling fountains or stagnant ponds: An integrative model of creativity and innovation implementation in work groups. ***Applied Psychology: An International Review,*** 51: 355–387.

Wilson, J. M., Goodman, P. S., & Cronin, M. A. 2007. Group learning. ***Academy of Management Review,*** 32: 1041–1059.

Wood, R. E. 1986. Task complexity: Definition of the construct. ***Organizational Behavior and Human Decision Processes,*** 37: 60–82.

Wright, T. 1936. Factors affecting the cost of airplanes. ***Journal of Aeronautical Science,*** 3: 122–128.

Xiao, Y., Hunter, W., Mackenzie, C., Jefferies, N., & the Lotas Group. 1996. Task complexity in emergency medical care and its implications for team coordination. ***Human Factors,*** 38: 636–645.

APPENDIX

Construct Validity of the Task Complexity Variable

To further support the construct validity of the task complexity measure and, in particular, to ensure that it tapped the two main aspects of surgical complexity—the complexity of a patient's condition and the complexity of the actual procedure performed—we conducted additional analyses. First, we tested for convergent validity between our single-item metric and patients' ASA scores, a standard surgical measure of patients' preoperative physical status. We found a moderate, positive correlation ($r = .29$, $p < .001$). Second, we asked two independent anesthesiologists who participate in many different surgeries conducted in different wards and who are employed by different hospitals to rate the complexity of each different surgery on the basis of both the procedure itself and the main diagnosis requiring surgical intervention. The anesthesiologists were asked to rate the complexity of the surgeries "as if they were all being conducted on the average patient." For each surgery, the surgeons were asked to respond to six items derived from Wood's (1986) definition of complexity. Two items addressed each of the three main complexity dimensions noted by Wood: component complexity, coordinative complexity (i.e., interdependence), and dynamic complexity. The interclass correlation for the agreement between judges was .64. Our single-item complexity measure correlated at .60 ($p < .01$) with this six-item, procedure-focused complexity measure, offering further evidence of our measure's construct validity. We chose to test our moderation hypotheses on the basis of our single-item measure of the complexity of each particular surgery rather than this more generic, procedure-focused measure because the former takes into account the interaction between patient condition and procedure and is thus likely to offer a greater degree of sensitivity with regard to the complexity actually experienced by the participants in the particular surgeries assessed.

Dana R. Vashdi (dvashdi@poli.haifa.ac.il) is a lecturer in the Department of Public Administration and Policy at the University of Haifa. Her research focuses on teamwork and team learning as well as on employee well-being, organizational climate, and health care policy. She received her Ph.D. in behavioral and management sciences from the Technion's Faculty of Industrial Engineering & Management in 2007.

Peter A. Bamberger (peterb@post.tau.ac.il) is professor of organizational behavior and human resource management at the Recanati Graduate School of Business Administration, Tel Aviv University, and senior research scholar at the School of Industrial and Labor Relations, Cornell University. Current research interests include employee helping and help seeking, team processes, employee well-being, and compensation strategy. He received his Ph.D. in organizational behavior from Cornell University in 1990.

Miriam Erez (merez@ie.technion.ac.il) is professor emeritus of organizational psychology, Faculty of Industrial Engineering & Management, Technion, Israel. Her research is in three major areas: (a) work motivation at the individual and team levels; (b) innovation at work; (c) cross-cultural organizational behavior. She received her Ph.D. in behavioral and management sciences from the Technion's Faculty of Industrial Engineering & Management in 1972.