AN INVESTIGATION OF THE EFFECT OF AFTER-ACTION REVIEWS ON TEAMS' PERFORMANCE-EFFICACY RELATIONSHIPS

A Dissertation

by

IRA ANTHONY SCHURIG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2012

Major Subject: Psychology

An Investigation of the Effect of After-Action Reviews

On Teams' Performance-Efficacy Relationships

AN INVESTIGATION OF THE EFFECT OF AFTER-ACTION REVIEWS ON TEAMS' PERFORMANCE-EFFICACY RELATIONSHIPS

A Dissertation

by

IRA ANTHONY SCHURIG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Winfred Arthur, Jr.
Committee Members,	Mindy E. Bergman
	Stephanie C. Payne
	Richard W. Woodman
Head of Department,	Ludy T. Benjamin, Jr.

May 2012

Major Subject: Psychology

ABSTRACT

An Investigation of the Effect of After-Action Reviews

on Teams' Performance-Efficacy Relationships. (May 2012)

Ira Anthony Schurig, B.B.S., Hardin-Simmons University;

M.A., Saint Mary's University

Chair of Advisory Committee: Dr. Winfred Arthur, Jr.

Performance and efficacy are reciprocally causal; however, the effect of performance on subsequent perceptions of efficacy has received little attention, especially in the context of team training. In addition, the moderating effect of feedback accuracy on the relationship between team performance and team-efficacy is largely unexplored. As such, the objective of the present study was to investigate the relationship between team performance and team-efficacy in the context of after-action reviews (AARs). Specifically, this study examined the conjoint influence of (a) the accuracy of performance feedback available to trainees during AARs, and (b) time on the predictive validity of team performance on team-efficacy. Data were obtained from 492 undergraduate students assigned to 123 teams in a 5 hr team training protocol using a 3 (training condition: non-AAR, versus subjective AAR, versus objective AAR) × 3 (sessions) repeated measures design.

Contrary to the first set of hypotheses, the positive relationship between performance and efficacy was strongest for teams trained without AARs and weakest for

teams trained using subjective AARs. Although team-efficacy was predicted more strongly by more proximal team performance than by more distal team performance, this pattern of results was found only for teams trained either without AARs or with objective AARs. The predictive validity of performance on efficacy decreased as performance episodes became more proximal among teams trained using subjective AARs. Finally, within-team agreement of team-efficacy ratings decreased over time for teams that engaged in AARs and remained constant over time for teams that did not engage in AARs.

The theoretical and practical implications of these findings are discussed. It is anticipated that this research will provide insight into the roles of feedback accuracy and time in the performance-efficacy relationship and provide guidance to researchers and practitioners in effectively integrating AAR design characteristics into team training environments.

ACKNOWLEDGEMENTS

First, I would like acknowledge Mindy Bergman, Stephanie Payne, and Dick Woodman for serving on my committee. I truly appreciate your time and guidance and it has been a pleasure working with each of you.

In addition, I'd like to thank everyone in the lab: Steven Jarrett, Ryan Glaze, Andrew Naber, Gonzalo Munoz, Jennifer McDonald, Olabisi Asenuga, and Craig White. I couldn't have written this dissertation without all your help.

I also want to express my sincerest gratitude and appreciation to my committee chair, Winfred Arthur, for his support, guidance, and friendship. I came to Texas A&M to work with you, and it's been a privilege working with and learning from you.

Finally, and most importantly…thank you, Margaret, Mycasslyn, Amanda, and Preston. I couldn't have done this without you.

DISCLAIMER

The views expressed in this work are those of the author and do not reflect the official policy or position of the United States Air Force, the Department of Defense, or the U.S. Government.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The objective of the current study was to investigate the relationship between team performance and team-efficacy in the context of after-action reviews (AARs). Bandura (1977, 2012) asserts that self-efficacy is a focal determinant of behavior because it affects behavior both directly and through its effects on other determinants of behavior, such as goals. In contrast to Bandura's assertion that efficacy is the primary predictor of performance, recent work has provided strong evidence that past performance is a better predictor of future performance than self-efficacy and that self-efficacy explains very little variance in future performance when examined longitudinally after controlling for past performance (e.g., Arthur, Bell, & Edwards, 2007; Heggestad & Kanfer, 2005; Judge, Jackson, Shaw, Scott, & Rich, 2007; Richard, Diefendorf, & Martin, 2006). Furthermore, it has also been demonstrated that within specific boundary conditions, self-efficacy can actually show a negative relationship with future performance when examined within-individuals (e.g., Vancouver, Thompson, Tischner, & Putka, 2002; Vancouver, Thompson, & Williams, 2001). One explanation for the negative effect of self-efficacy on performance is the accuracy of performance feedback (Schmidt & DeShon, 2010); however, this finding has been limited to the individual level of analysis. Although the effect of efficacy on performance remains unclear, what is less clear is the effect of feedback accuracy on the performance-efficacy relationship in the context of teams, specifically, team training.

Kozlowski and Ilgen (2006) stated that "Teams are central and vital to everything

---

This dissertation follows the style of *Journal of Applied Psychology*.

we do in modern life" (p. 78) and as the use of teams in organizations becomes more prevalent, the need to understand and improve team functioning across a team's lifespan has increased. In addition, as organizations have increasingly structured work around the use of teams (Devine, Clayton, Philips, Dunford, & Melner, 1999), they have concomitantly sought to develop and implement strategies for enhancing the effectiveness of such work teams.

Bandura's (1977) social cognitive theory has long been a dominant framework for analyzing behavior. In extending social cognitive theory from the individual to the group level of analysis, Bandura (1982) introduced the term "collective efficacy"—also referred to as team-efficacy—and posited that team-efficacy has a positive causal effect on team performance. However, teams frequently function in temporal cycles of goal-directed activity consisting of multiple performance episodes (Marks, Mathieu, & Zaccaro, 2001). In such temporal cycles, team members will evaluate the team's prior performance in order to assess how well prepared they are for future team functioning. When team functioning is viewed as a series of temporal cycles, it becomes apparent that understanding both the predictive influence of team-efficacy on team performance *and* the predictive influence of team performance on team-efficacy are critical for gaining additional insight into team functioning. Considerable evidence supports the predictive influence of team-efficacy on team performance (e.g., Bandura, 2000; Gibson, Randel, & Earley, 2000; Hodges & Carron, 1992; Little & Madigan, 1995; Zaccaro, Blair, Peterson, & Zazanis, 1995). However, in contrast to the large body of work which has

examined performance as a dependent variable, very little empirical research addresses the question of whether team performance predicts team-efficacy.

Lindsley, Brass, and Thomas (1995) referred to the temporal cyclical relationship between performance and efficacy as a performance-efficacy spiral and proposed that a key to self-correcting adjustment—and hence, more effective team functioning—is timely and accurate performance feedback. In the context of accurate feedback, a team's assessment of its efficacy will be more congruent with its actual past performance of the task and as such, can be used more effectively by the team in allocating resources in pursuit of successful task performance. For example, a team that performs poorly and then consequently assesses its team-efficacy as low should persist longer and devote more effort to subsequent task performance—resulting in improved performance. In contrast, a team that performs poorly and does not recognize its team-efficacy as low may not persist longer and devote more effort to subsequent task performance—resulting in further poor performance.

One particularly powerful means by which teams gain knowledge and understanding of prior team performance episodes is through an after-action review (AAR)—a feedback-based learning procedure in which trainees analyze behavior and evaluate contributions of behaviors to performance outcomes (Ellis & Davidi, 2005). Although corporate and military organizations embrace the use of AARs, there are very few empirical investigations of AARs in team training contexts, and still fewer investigations of the effect of AARs on emergent team processes.

Although feedback interventions have variable effects on performance, research has generally supported the notion that feedback—especially feedback that is perceived as unbiased or objective—can enhance task performance by directing attention to aspects of the task on which feedback is provided, which results in error reduction and goal setting (Ilgen, Fisher, & Taylor, 1979; Kluger & DeNisi, 1996). Empirical evidence suggests that the accuracy of performance feedback influences the efficacy-performance relationship at the individual level (Schmidt & DeShon, 2010; Shea & Howell, 2000), but—as best as can be ascertained—the present study is the first to examine the effect of the accuracy of performance feedback on the performance-efficacy relationship at the team level. As such, the major objective of this dissertation is to attempt to explore the role of the accuracy of performance feedback on team-efficacy—that is, on teams' ability to accurately calibrate performance and assessments of ability.

Team-Efficacy

Bandura (1997) defined team-efficacy as a team's "shared belief in its conjoint capabilities to organize and execute the courses of action required to produce given levels of attainments" (p. 477). Researchers have described team-efficacy as a task-specific motivational state that reflects team members' shared belief regarding aspects of their capabilities and tasks (Chen & Kanfer, 2006) and have further theorized that team-efficacy is an "emergent state" that arises from both individual states and beliefs and from interactions among team members (Marks et al., 2001). Bandura (1997) noted that although self-efficacy and team-efficacy differed in their unit of agency, both forms of efficacy were task-specific and originated from similar sources, served similar functions,

and operated according to similar mechanisms. Thus, according to Bandura (1997), team-efficacy should arise from any or all of the following factors: (a) *enactive mastery* experiences that indicate team capabilities on a specific task or set of tasks; (b) *vicarious experiences* through which team members acquire competencies and compare capabilities to others; (c) *verbal persuasion and social influence* through which team members gain information about their capabilities from an external perspective; and (d) *physiological and affective states* from which team members assess their capabilities, strength, and vulnerability to dysfunction.

Although team-efficacy and self-efficacy arise from similar sources, serve similar functions, and operate according to similar mechanisms, Bandura (2000) noted that team-efficacy was not simply the sum of team members' self-efficacy beliefs. Furthermore, Bandura (1997) noted that effective team functioning likely involves more complex reciprocal paths of social influence than does individual functioning and also posited that as members of a team attempt to coordinate their individual actions, they are likely to be influenced by their teammates' beliefs, motivations, and behaviors. Based on this more complex interplay of social influences, the construct of team-efficacy can and should be distinguished from self-efficacy.

Mischel and Northcraft (1997) also referenced this distinction between team- and self-efficacy when they suggested that the question of "can *we* do this task?" is different from the question of "can *I* do this task?" This distinction between the team- and individual-level ratings of efficacy is also echoed in discussions of how best to operationalize team-efficacy. That is, although some constructs—such as efficacy—may

be homologous at the individual and team levels (Kozlowski & Klein, 2000), the operationalization of efficacy may be qualitatively different at the two levels of measurement (Arthur et al., 2007).

In general, two aggregation-based approaches to the measurement of team-efficacy have been discussed in the literature: (a) aggregating team members' ratings of their individual capabilities for a particular task, and (b) aggregating team members' appraisals of their team's capability for a particular task (Bandura, 1997, 2000). These two operationalizations are analogous to Chan's (1998) additive and referent-shift consensus composition models. In an additive composition model, a higher-level construct (e.g., team-efficacy) is simply a summation of team members' perceptions of an individual-level construct (e.g., self-efficacy); thus, the focus of assessment is the lower-level construct (self-efficacy). In such models, variance in the lower-level construct is of no theoretical concern and the validity of the additive index (i.e., the mean) serves as empirical support for the composition (Chan, 1998). In contrast, the focus of assessment in a referent-shift consensus model is the higher-level construct (team-efficacy). The method of composition in referent-shift consensus models is also a simple summation; however, within-group consensus—as indexed by within-group agreement of individual team members' scores—serves as empirical support for the composition.

Although team-efficacy can be operationalized using either the additive or referent-shift consensus composition models, it is thought that the referent-shift consensus composition, which presumably takes into account the dynamic processes that

occur in a team context, more accurately measure team-efficacy (Gist, 1987) and better predict team performance, especially in tasks requiring a high degree of interdependency (Bandura, 2000). Several primary studies (e.g., Arthur et al., 2007; Katz-Navon & Erez, 2005) and a meta-analysis (Gully, Incalcaterra, Joshi, & Beaubien, 2002) have provided empirical support for this position.

The relationship between team performance and team-efficacy has traditionally been explained from a self-regulation perspective by researchers (e.g., Chen, Thomas, & Wallace, 2005; DeShon, Kozlowski, Schmidt, Milnet, & Weichmann, 2004) who have extended Bandura's (1997) social cognitive theory to the team level. Social cognitive theory posits a triadic reciprocal causal model in which behaviors, interpersonal factors (in the form of cognitive and affective events), and the external environment all influence each other in a dynamic manner (Bandura, 1997). This theory is based on a perspective of human agency in which "people function as anticipative, purposive, and self-evaluating proactive regulators of their motivation and actions" (Bandura & Locke, 2003, p. 87). Bandura (1997) suggested that team-efficacy influences the goals a team sets for itself, how much effort the team expends to achieve these goals, and persistence in the team's expenditure of such effort. Empirical research has largely provided support for Bandura's position. Prussia and Kinicki (1996) demonstrated that collective efficacy was related to collective goals and performance, and Chen, Kanfer, DeShon, Mathieu, and Kozlowski (2009) found that team-efficacy positively predicted team action processes and subsequent team performance and was itself predicted by prior team performance. In their meta-analysis, Gully et al. (2002) reported an overall

sample-weighted mean correlation between team-efficacy and team performance of 0.41, and further found that this relationship was stronger for teams that performed tasks requiring higher levels of interdependence (estimated true score $r = 0.45$).

In extending social cognitive theory to explain team functioning, several researchers (e.g., Chen et al., 2005; DeShon et al., 2004; Gibson, 2001) have found evidence of homology between both individual-level performance and self-efficacy and team-level performance and team-efficacy, suggesting that the relationships between performance and efficacy at the individual and team levels are subject to similar dynamics. In addition, meta-analytic findings obtained by Gully et al. (2002) on the relationship between team-efficacy and team performance (estimated true score $r = 0.41$) are similar in magnitude to Stajkovic and Luthans' (1998) meta-analytic findings on the relationship between self-efficacy and individual performance (estimated true score $r = 0.38$). Similarities in the magnitude of these meta-analytic findings, the evidence of homology between the individual and team level, and the correspondence in the conceptual definitions of self- and team-efficacy all provide evidence that the relationship between team performance and team-efficacy can reasonably be explained from a social cognitive perspective.

The reciprocal nature of the social cognitive model, in conjunction with Bandura's (1997) assertion that efficacy beliefs are derived from enactive mastery experiences, implies that a team's performance experience influences subsequent perceptions of team-efficacy. It should also be noted that Gully et al. (2002) refrain from making any causal inferences based on their meta-analytic findings on the

relationship between team-efficacy and team performance. They explicitly state that "it is difficult to believe that teams will engage in an activity if they feel it is impossible for members to accomplish anything. However, performance is equally likely to influence subsequent efficacy" (p. 828).

Causality in the efficacy-performance relationship at the individual level has been debated since Bandura introduced the construct (Borkovec, 1978; Corrigan, 1990; Hawkins, 1992; Lee, 1989). For example, some (e.g., Bandura, 1982; Earley & Lituchy, 1991) have argued that self-efficacy is a cause of behavior; in contrast, others (e.g., Borkovec, 1978; Hawkins, 1992) have posited that self-efficacy is only a predictor of behavior. According to Bandura (1997), efficacy perceptions are based on success experiences and the information those experiences convey, such that successful performances raise individuals' efficacy beliefs and failures lower them. Borkovec (1978) argued that because behavioral change could be attributed to learning without making reference to unobservable cognitions, self-efficacy is better viewed as a consequence rather than a cause of behavioral change. Bandura (1982) seems to have acknowledged this position by stating that "performance mastery, in turn, can boost perceived self-efficacy in a mutually enhancing process" (p. 128). Hawkins (1992) later argued that in Bandura's (1977) "enactive mastery" treatment of snake phobia, it is actually direct experience—not increases in self-efficacy—which account for decreases in snake phobia. This direct experience is also referred to by Bandura as enactive mastery and is one of the four main sources of self-efficacy (1997). In addition, Hawkins (1992) further argues that with reference to learning principles and Bandura's

snake phobia treatment, the enactive mastery treatment—not self-efficacy as Bandura (1977) suggested—should be the salient independent variable. Thus, Hawkins holds the view that self-efficacy should have an origin and previous behavior is undoubtedly critical to this origin. Therefore, taken together and extended to the team level, the arguments of Borkovec and Hawkins provide a reasonable foundation upon which to take the position that team-efficacy is a consequence of team performance. Thus, it is hypothesized that:

Hypothesis 1a: Team performance will positively predict team-efficacy.

In addition to the rational arguments for performance as a cause of efficacy beliefs, recent empirical research (e.g., Arthur et al., 2007; Heggestad & Kanfer, 2005; Judge et al., 2007; Richard et al., 2006; Shea & Howell, 2000; Vancouver et al., 2002; Vancouver et al., 2001) has also shed light on the need to examine the effect of performance on efficacy. This research, based on Powers' (1973) control theory, has provided evidence that the positive relationship between self-efficacy and performance is more a product of the effect of past performance on self-efficacy than the effect of self-efficacy on subsequent performance.

Powers' (1973) control theory of human functioning—similar to Bandura's social-cognitive theory—relies on goals as a key motivational construct. However, control theory is based on a negative feedback loop in which deviation from a desired state triggers actions to drive a system toward the desired state. For example, a student might have a goal of scoring a 90 on an upcoming exam and may also have a perception of his or her ability to score a 90. According to control theory, the student will be

motivated to reduce the discrepancy between the desired level of preparedness for the exam and the current state of preparedness for the exam. In this case, if the student believes he or she can currently score an 80, yet wants to score a 90, then the student will allocate resources (i.e., study) until the discrepancy is eliminated or the student runs out of study time. A critical aspect to note is that the perception of preparedness is based on a subjective assessment—rather than direct knowledge—of preparedness (Koriat, 1997) and one measure of that subjective assessment is a perception of efficacy (Vancouver & Kendall, 2006).

Recently, control theory has been invoked to explain findings indicating that efficacy does not always facilitate performance, and in some cases may *negatively* influence performance. Some researchers argue that high self-efficacy may lead to overly optimistic interpretations of performance, resulting in the perception that one is more prepared to attain a given performance goal than one actually is (Powers, 1991; Vancouver et al., 2002; Vancouver et al., 2001). This perceived smaller discrepancy between a desired goal and a current level of preparedness is argued to lead to reduced allocation of resources such as time and effort (Vancouver & Kendall, 2006; Vancouver, More, & Yoder, 2008).

Bandura (1997) stated that "efficacy beliefs cannot operate as a regulative influence in an informational vacuum" (p. 66) and suggested that a lack of accurate feedback compromises the benefits of efficacy. Researchers attempting to explain the negative effect of self-efficacy on performance (Powers, 1991; Vancouver et al., 2002; Vancouver et al., 2001) have theorized that ambiguity concerning the true status of one's

performance may be a critical component of this negative efficacy effect. In response to Bandura and Locke's (2003) assertion that a test of such a theory would require comparing the effects of self-efficacy on performance under varying levels of feedback ambiguity, Schmidt and DeShon (2010) found that the relationship between self-efficacy and performance was moderated by level of feedback ambiguity. Specifically, they found that under conditions of low ambiguity feedback, (i.e., participants were aware of the true status of their performance), self-efficacy positively predicted level of effort, which then positively predicted performance on an anagram task; thus, self-efficacy positively predicted performance in a low ambiguity feedback context. In contrast, under conditions of high ambiguity (i.e., participants were unaware of the true status of their performance), self-efficacy negatively predicted level of effort, which then positively predicted performance; thus, self-efficacy negatively predicted performance in a low ambiguity feedback context. In addition, Schmidt and DeShon also performed an auxiliary analysis in which they found that performance positively predicted subsequent perceptions of self-efficacy. They found that the relationship between prior performance and subsequent efficacy did not vary as a function of the level of feedback ambiguity; however, they investigated this relationship at the individual level of analysis. Therefore, it remains unclear whether level of feedback ambiguity influences how accurately individuals working as a team judge the team's level of preparedness.

The AAR

The AAR was originally developed by the military decades ago, and although several names exist (e.g., after-event review, debrief, hotwash), it has remained

relatively unchanged since its inception. At the broadest level, the AAR is an approach to training that turns a recent event into a learning opportunity by allowing participants to systematically review and discuss a task or event of interest. The U.S. Army (1993) defines the AAR as "a professional discussion of an event, focused on performance standards, that enables soldiers to discover for themselves what happened, why it happened, and how to sustain strengths and improve on weaknesses" (p. 1). Ellis and Davidi (2005) define the AAR as "an organizational learning procedure that gives learners an opportunity to systematically analyze their behavior and to be able to evaluate the contributions of its various components to performance outcomes" (p. 857). Thus, an AAR allows trainees to systematically review their performance on a recently completed task or event.

As illustrated in Figure 1 (Villado, 2008), trainees in an AAR seek answers to the following questions concerning a training or performance episode: What was the intended outcome? What was the actual outcome? What specific actions and behaviors contributed to meeting the intended outcome? What specific actions and behaviors detracted from meeting the intended outcome? What is the intended future outcome? What actions will increase the likelihood of meeting the intended future outcome?

Villado (2008; Villado & Arthur, 2011) partitioned the AAR into five phases and proposed that existing psychologically-based theories provided guidance for a more thorough understanding of the AAR. Trainees receive *feedback* in the first two phases of the AAR as they review the intended objective and the actual outcome of the previous performance episode. This feedback serves as a basis for performance

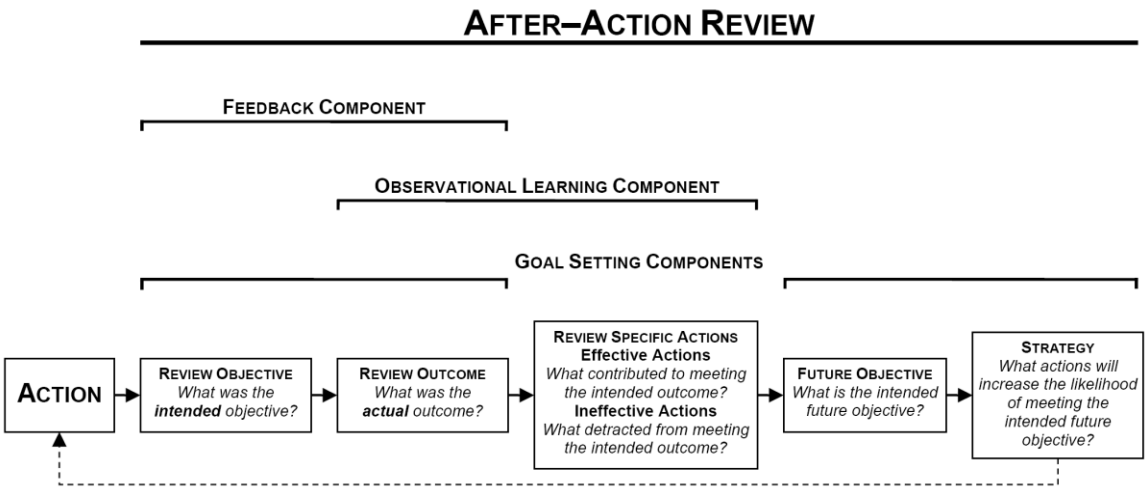


Figure 1. Primary phases of the after-action review and associated psychological theory. Adapted from "The after-action-review training approach: An integrative framework and empirical investigation," by A. J. Villado, 2008, *Proquest Dissertations and Theses*, 3333780, p. 17. Copyright 2008 by Anton J. Villado with permission.

evaluation. In the third phase, trainees *observe behaviors* performed (by themselves and/or other team members) and learn through observation and reflection. Finally, the final two phases provide trainees with an opportunity to *set goals* and develop strategies for achieving these goals. Thus, Villado and Arthur propose that an integration of the feedback, observational learning, and goal-setting literatures provides a theoretical explanation of the effectiveness of AARs as a training method.

In addition to Villado and Arthur's (2011) theoretical explanation for the effectiveness of AARs, Ellis and Davidi (2005) proposed that an AAR facilitated three functions critical to learning—self-explanation, data verification, and feedback. Self-explanation is described as an active process of gathering, analyzing, and integrating data which serves to direct learners to reflect on past performance and to encourage

integration of new information with existing knowledge (Chi, de Leeuw, Chiu, & Lavancher, 1994). In addition, AARs provide an opportunity for participants to validate their own perceptions against external sources of information (e.g., team members' perceptions or objective performance feedback). This data verification process serves to minimize the effects of cognitive errors such as hindsight bias—the influence that knowledge of outcomes has on views of past experiences (Fischhoff, 1982)—and confirmation bias—the tendency to overlook information not compatible with a priori hypotheses (Brehmer, 1980). Confronting different perceptions of the same data enables participants in AARs to integrate external sources of information into their own mental models (Ellis & Davidi, 2005). Finally, the feedback that participants receive during AARs is focused not only on performance outcomes, but also on the process of task performance, metacognitive knowledge, beliefs about self and task, and cognitive strategies (Alexander, Schallert, & Hare, 1991). Ellis and Davidi (2005) also suggest that functional validity feedback—a mechanism for learning improvement which helps learners understand the gaps between estimates of their achievements and their actual performance (Balzer, Doherty, & O' Connor, 1989)—is a by-product of the AAR process.

Although the structure and conduct of AARs have changed since its initial use as a training intervention, the more substantial changes to the AAR have involved the integration of various technological advances intended to facilitate the conduct of the review (e.g., video and other recording equipment to objectively document performance, integration of recording and rating tools into simulators and simulation software) and

provide more accurate performance feedback to trainees. Therefore, present day AARs may differ greatly from those of the past in terms of fidelity and objectivity because of technological advances in recording, playback, and evaluation systems. However, the purpose of the AAR remains unchanged—to systematically review trainees' performance on a recent task or event in order to create a learning opportunity with the aim of improving subsequent performance.

AARs conducted during both individual and team training allow for learning from prior performance, but there are several aspects of an AAR that makes it especially appropriate for team training. First, vicarious learning occurs during AARs. During team AARs, individual team members learn not only from their own individual performances, but they also learn from observing and discussing the performance of others (Bandura, 1986). Second, team members share information during the AAR process. In their meta-analysis of the effects of information sharing on team performance, Mesmer-Magnus and DeChurch (2009) found that team information sharing was positively related to team performance. During team training, an AAR provides a venue for information sharing among team members as they review a past training event. Because each team member provides a unique perspective and experience of the event, a discussion of what happened naturally promotes information sharing. Third, an AAR provides a forum for interaction among team members. During an AAR, team members *collectively* review prior performance and set goals. In doing so, they move toward a shared belief in their team's efficacy (Villado, 2008).

State of the AAR literature. Although the scope of the AAR-based literature is limited, several recent primary studies provide empirical support for the position that AARs are an effective training method (e.g., Alexander, Kepner, & Tregoe, 1962; Ellis & Davidi, 2005; Ellis, Ganzach, Castle, & Sekely, 2010; Ellis, Mendel, & Nir, 2006; Smith-Jentsch, Cannon-Bowers, Tannenbaum, & Salas, 2008).

Trainees who participate in AARs have higher post-training performance (Alexander et al., 1962; Ellis et al., 2010; Ellis et al., 2006; Villado, 2008) and non-performance outcomes (Ellis et al., 2010) than trainees who do not participate in AARs. Ellis et al. (2006) reported standardized mean differences between AAR and non-AAR post-training performance ranging from 0.03 to 1.02. More recently, in a meta-analysis of 15 independent samples from seven studies that met their inclusion criteria, Schurig, Jarrett, Arthur, Glaze, and Schurig (2011) obtained an overall sample-weighted mean *d* of 1.12 for AARs, and a sample-weighted mean *d* of 0.75 for AARs in team-training contexts, indicating they are an effective training method.

AARs typically result in faster performance improvement than control training conditions (e.g., Ellis et al., 2006; Villado, 2008). However, Alexander et al. (1962) identified a situation in which AARs may not have resulted in performance gains as much as AARs may have prevented performance declines. Specifically, teams trained using AARs not only had better overall performance scores at the end of training, but on some dimensions of performance, the post-test performance of teams in the control group was worse than their pre-test performance. Alexander et al. found that when teams had no knowledge of results, teams trained using AARs demonstrated increases in

performance, whereas those trained without AARs did not show performance increases. However, when teams had clear knowledge of results, the difference in performance between teams trained with and without AARs was smaller. Thus, the effectiveness of the AAR was moderated by the extent to which trainees were aware of the results of their actions, with AARs being more effective for tasks where knowledge of results was lacking.

The specific aspects of a performance experience on which trainees focus during an AAR also appears to moderate its effectiveness. Ellis and Davidi (2005) found that reviewing both successful and unsuccessful performance experiences enhances trainees' learning. Specifically, trainees who reviewed both successes and failures during AARs generated richer mental models and had greater performance increases than trainees who only reviewed failures during AARs. Ellis and Davidi concluded that although trainees, trainers, and organizations may focus on error reduction, reviewing both successful and unsuccessful performance enhances trainees' learning and task conceptualization; they posited that reviewing failures during AARs motivated epistemic processes such as hypothesis generation and information gathering more than did reviewing successes.

Ellis et al. (2006) extended this research to include AARs that focused on teams' (a) successes only, (b) failures only, or (c) successes and failures. In a laboratory-based market simulation task, they found that after a successful performance, only a failure-focused AAR resulted in subsequent performance improvement, but after an unsuccessful performance, all types of AARs resulted in subsequent performance improvement. Ellis et al. speculated that directing learners to gather and elaborate on

errors caused learners to question the appropriateness of their knowledge, which subsequently boosted their motivation to test, update, and integrate it into future task performance.

Smith-Jentsch et al. (2008) investigated the effect of guided team self-correction on shared cognitions, team processes, and team performance. Guided team self-correction is a team debriefing strategy in which participants are responsible for diagnosing and solving problems with guidance regarding which topics to discuss and how to constructively discuss these topics (Smith-Jentsch, Zeisig, McPherson, & Acton, 1998). Using an expert mental model as a guide, Smith-Jentsch et al. (2008) found that systematically reviewing performance using the expert mental model as a review framework facilitated mental model accuracy, team processes, and performance in contrast to a less participative chronological review, presumably because chronological debriefs are thought to lead participants to develop mental models that (a) are organized in terms of concrete task features and (b) result in decreased generalization of lessons learned to novel situations (Smith-Jentsch et al., 2008).

Ellis et al. (2010) investigated the use of filmed AARs (with a facilitator and a model participating in a mock AAR). Such filmed AARs afford a trainer more control over the AAR process than does a typical AAR in which participants review their own performance. For example, in a filmed AAR, the content reviewed during the AAR is selected a priori. Thus, training is standardized when such filmed AARs are used because AARs are identical across all trainees. In this study, trainees watched a previously recorded AAR—in which an experimenter guided a model through a

discussion and review of performance—rather than participating in a discussion and review of their actual performance. Thus, the filmed AAR not only provided trainees with feedback on the task (albeit feedback of another trainee's performance), but also served as a behavioral model for the conduct of an AAR. Watching a filmed AAR was also thought to maintain trainees' psychological safety because they would not feel threatened or blamed for errors of the filmed trainee, but would still be able to learn from errors. Ellis et al. (2010) found that both types of AARs—personal and filmed—resulted in performance improvement, but they found no significant difference in performance improvement between the personal and filmed AARs.

Objective versus subjective AARs. As originally designed and implemented, AARs relied on the ability of trainees and/or a facilitator to recall and evaluate behaviors or critical incidents that occurred during a performance experience. However, technological advancements have allowed for the collection of objective, high-fidelity performance data that can be reviewed during AARs. In fact, the vast majority of the AAR literature—and considerable resources such as time and funding—has been concerned with developing, incorporating, and implementing systems that provide trainees with objective performance data during AARs in hopes of providing the most timely and accurate feedback possible. As such, one can draw a distinction between "subjective" AARs that rely exclusively on trainees' and facilitators' ability to recall and evaluate performance data and "objective" AARs that rely on a variety of veridical data sources—such as video or audio recordings, flight data, or objective performance logs—to facilitate recall and evaluation of performance experiences. Thus, a critical difference

between objective and subjective AARs is that when a team reviews a previous performance episode (i.e., the “Review Outcome” phase in Figure 1), teams in an objective AAR have an opportunity to verify ideas about what happened against a true account of what actually happened (e.g., a video recording of the performance episode).

In spite of the resources that have been devoted to collecting and packaging objective performance data for review during AARs, it appears that practice has outpaced science in that empirical investigations examining the effect of objective review systems are lacking. Ellis et al. (2010) recently investigated the effect of having trainees watch other trainees participate in an AAR, but these filmed AARs did not allow trainees to review objective data regarding their own performance. Savoldelli et al. (2006) investigated the value of a debriefing process in the context of simulation-based education. They found that when anaesthesiologist residents received no feedback concerning performance, their skills did not improve. The provision of both oral only and video-supplemented oral feedback resulted in skill improvement; however, they found no difference in performance levels between oral only and video-supplemented oral feedback conditions. It should be noted that the debriefings conducted by Savoldelli et al. all involved objective data. That is, both feedback conditions involved debriefings facilitated by instructors who provided data to learners—either orally or orally with a video supplement—and encouraged learners to reflect on the feedback. In contrast, Villado and Arthur (2011) and Arthur et al. (2011) have investigated the comparative effectiveness of objective and subjective AARs. Villado and Arthur (2011) found that both subjective AARs (i.e., AARs in which trainees relied on their own memories to

recall the intended and actual outcomes of the most recently completed training event) and objective AARs (i.e., AARs in which trainees reviewed the progress of their most recently completed training event using a video replay of their actual performance) resulted in greater performance and team-efficacy gains than a control condition. However, objective AARs did not result in higher performance or team-efficacy than subjective AARs. Arthur et al. (2011) found similar results; both objective and subjective AARs resulted in higher performance and team-efficacy than a non-AAR control condition. However, they found these results only for geographically co-located teams (i.e., all team members were physically in the same room), not for geographically distributed teams (i.e., team members were physically located in different buildings).

Despite the limited empirical research investigating the effect of objectivity during AARs, the performance appraisal and assessment center literatures have examined the effect of objectivity on the comprehensiveness and accuracy of evaluations and assessments (DeNisi, Robbins, & Cafferty, 1989; Sturman, Cheramie, & Cashen, 2005) and offer some insight into the use of objective AARs.

Performance appraisals made both with and without memory aids (e.g., diaries and notes) demonstrate similar levels of rating accuracy in terms of assessments of the overall performance of a target (Middendorf & Macan, 2002; Ryan et al., 1995; Sanchez & De La Torre, 1996; Woehr & Feldman, 1993). Similar results have been reported in the assessment center literature. Specifically, assessors making ratings during an exercise and assessors viewing videotaped recordings of the same exercise made equally accurate ratings (Ryan et al., 1995). In contrast to ratings made based on memory aids,

research has also demonstrated that performance ratings made solely from memory have less recall accuracy (Middendorf & Macan, 2002; Ryan et al., 1995; Sanchez & De La Torre, 1996; Woehr & Feldman, 1993). DeNisi et al. (1989) found that raters who relied on memory when evaluating a target recalled fewer incidents and made more recall errors than those who were allowed to use a diary-like memory aid. This finding is particularly noteworthy given the relatively short time interval (i.e., often only several minutes) between observing and rating performance. Similar research has demonstrated that recall errors become more pronounced as the time between observing and rating performance increases (DeNisi et al., 1989; Murphy & Balzer, 1986; Williams, DeNisi, Meglino, & Cafferty, 1986). Finally, Ryan et al. (1995) found that providing assessors with access to videotaped recordings of assessment center exercises resulted in greater observation quality and slightly better observation accuracy, but only if assessors were given the ability to control (e.g., pause and rewind) the videotaped recordings.

Taken together, it appears that raters are able to form general performance evaluations while observing behavior and are able to provide subsequent ratings based on those general evaluations (Murphy & Balzer, 1986; Woehr & Feldman, 1993). In contrast, when raters do not form general evaluations prior to providing a formal rating, raters base their ratings on memory (Murphy & Balzer, 1986; Woehr & Feldman, 1993). As the time between performance observation and evaluation increases, ratings are less influenced by details of the performance and are more influenced by the general impression formed about the performance (Murphy & Balzer, 1986). Thus, recall accuracy is not as critical for accurate performance appraisals as one would expect; if

raters are able to form accurate internal evaluations prior to an appraisal and are able to access those internal evaluations, then they merely need to draw on those evaluations when rating a target (Murphy & Balzer, 1986; Woehr & Feldman, 1993).

Although it may be feasible for experienced raters to accurately generate internal evaluations during a workday that can be used for subsequent performance appraisals, it seems less likely that trainees would be able to do so during training. Trainees' ability to accurately identify, encode, and form internal evaluations of critical incidents is likely to be decreased due to the cognitive demands of learning new tasks while simultaneously attempting to attend to and evaluate performance. In addition, it is even less likely that trainees would be able to simultaneously identify, encode, and evaluate the performance of other trainees during a team task given the difficulty of noting and evaluating *their own* performance.

Precise recall of behaviors is needed for effective performance feedback (Murphy, 1991), and when reviewing performance, errors in recall may result in trainees omitting behaviors or critical incidents that affected performance. Errors in recall may also result in trainees including irrelevant or counterfactual behaviors in a performance review. Such recall errors—whether they result in deficiency or contamination—may diminish the effectiveness of an AAR. That is, when trainees are less able to generate memory aids for their own behavior and the behavior of teammates, objective review methods may enhance the effectiveness of AAR-based training.

In summary, objective AARs should provide trainees with more accurate performance feedback than do subjective AARs. Neither objective nor subjective AARs

provide a mechanism whereby trainees can generate more accurate internal evaluations of performance *during* performance. However, during a systematic review of performance, the veridical data available in an objective AAR—such as audio and video replay capability—provides trainees more accurate feedback about their prior performance than does the self-generated feedback available to trainees in a subjective AAR.

Performance and Efficacy Spirals

The relationship between performance and efficacy is cyclical. That is, team performance is thought to affect team-efficacy, which is thought to in turn affect subsequent team performance, and so on. Lindsley et al. (1995) termed these iterative loops "performance spirals." Such performance spirals can often become "deviation amplifying" (Henschel, 1976; Masuch, 1985; Weick, 1979). For instance, in deviation-amplifying loop, a decrease in performance causes a decrease in efficacy, which in turn causes a decrease in performance, and so on, in an amplifying relationship. Lindsley et al. (1995) further argue that the relationship between efficacy and performance over time is likely complicated and may be punctuated by corrections in either performance or efficacy. In addition, Lindsley et al. argue that performance and efficacy are so highly interdependent that to focus on a single variable or to determine unidirectional causality would obscure the potentially amplifying properties of such a spiral. Thus, it is the pattern of relationships that is the critical feature of an amplifying loop—one must change the *relationship* between performance and efficacy in order to change the loop. This change can be accomplished via efficacy beliefs that accurately reflect actual

performance. For example, if analysis of performance allows a team to make adjustments in future efforts that reverse a previous decrease in performance and efficacy, then the team's performance-efficacy spiral would be changed. Thus, the cyclical nature of the performance-efficacy relationship suggests three possible patterns: a downward spiral (e.g., decreasing performance and efficacy), an upward spiral (e.g., increasing performance and efficacy), or a self-correcting cycle (e.g., a decrease in both performance and efficacy causes an increase in either performance or efficacy as one learns from mistakes).

Although at face value upward spirals seem desirable and beneficial to team functioning, Lindsley et al. (1995) did not equate upward spirals with the positive motivational effects and performance increases that result from positive expectations, goal-setting, or other interventions aimed at increasing efficacy. Instead, they agreed with Gist (1987) that increasing efficacy without increasing learning would lead to overconfidence. The difficulty, Lindsley et al. (1995) note, is in distinguishing between the positive effects of confidence and the complacency of overconfidence. They propose that an upward spiral (defined as three consecutive increases in both efficacy and performance) should be positively related to complacency and overconfidence because continual success decreases the active experimentation necessary for improvement (March, 1976).

Lindsley et al. (1995) echo the beliefs of Sitkin (1992) and Weick (1979) in stating that self-correcting cycles are preferable to both upward and downward spirals, (which Lindsley et al. considered isomorphic) because "an increase in long-term

performance is not achieved without occasional failure and learning from one's mistakes" (p. 651). Thus, it is the congruency of efficacy beliefs with performance that is critical to team effectiveness. The more a team's efficacy belief reflects actual performance, the more effectively the team can mobilize effort and develop analytic strategies in support of future performance (Wood & Bandura, 1989).

The Role of AARs in the Performance-Efficacy Relationship

Lindsley et al. (1995) note that the key to self-correcting adjustment and avoiding deviation-amplifying spirals is acquiring timely and accurate performance feedback. Bandura (1997) echoed this sentiment in positing that a lack of timely and accurate feedback compromises the benefits of self-efficacy. Taken together, a critical aspect of the performance-efficacy loop is revealed—accurate performance feedback is necessary for teams to calibrate performance and efficacy. Lindsley et al. (1995) further note that simply having knowledge of success or failure on a previously performed task is not sufficient for the occurrence of self-correcting adjustment or avoidance of deviation-amplifying spirals. Instead, feedback must be accurate, timely, and specific in order to result in an understanding of the cause-and-effect relationships involved in performing a task (Ashford, 1989). According to Ashford, feedback draws attention to the matching-to-standard process, thereby enabling individuals to re-evaluate their ability to achieve a performance goal. Therefore, more accurate feedback should increase the congruence between a team's assessment of its capabilities (i.e., team-efficacy) and previous team performance because of the availability of accurate information about performance (Shea & Howell, 2000). Because both prior team performance and team-efficacy are

implicated in future team performance, a major contribution of this study to the extant team training literature is the investigation the role of the accuracy of feedback in the team performance-team-efficacy relationship.

Schmidt and DeShon (2010) found—albeit at the individual level—that feedback ambiguity moderated the relationship between efficacy and performance. They suggested that uncertainty regarding the true status of one's performance is an essential factor underlying the negative self-efficacy effect (Vancouver et al., 2002; Vancouver et al., 2001). Schmidt and DeShon subsequently found that self-efficacy positively predicted performance under conditions of low feedback ambiguity (i.e., when participants were aware of the true status of their performance) and negatively predicted performance under conditions of high feedback ambiguity (i.e., when participants were unaware of the true status of their performance). In addition, they found that performance predicted self-efficacy, but did not report whether this relationship was moderated by level of feedback ambiguity. Because efficacy beliefs are based on enactive mastery, it is not unreasonable to posit that accuracy of feedback should moderate how strongly performance predicts efficacy. For example, given equal levels of prior performance, a team receiving more accurate feedback about that performance should more accurately gauge its efficacy than a team receiving less accurate feedback about that performance. Shea and Howell (2000) also found—again at the individual level of analysis—that the presence of feedback was associated with self-correcting performance spirals. That is, they found that individuals who received feedback were more likely to make self-efficacy assessments that were more reflective of actual task

performance. Extending the theoretical foundations underlying this empirical work to the team level results in the inference that more accurate feedback results in team-efficacy beliefs that are more congruent with actual task team performance.

The foundation on which this supposition rests is that team-efficacy is an estimate of prior performance. When a team is unclear about its level of performance, it must be estimated. More accurate feedback results in more accurate estimates of performance, and more accurate estimates of team performance (and hence, team-efficacy) result in more appropriate actions taken in pursuit of goals (Powers, 1973, 1978; Vancouver et al., 2002; Vancouver et al., 2001).

Masuch (1985) posited that when the number of causal relationships required for successful task performance is large (i.e., the task is complex) or when the causal relationships required for successful task performance are unknown, ambiguous, or unpredictable (i.e., the task is uncertain), a full understanding of all the relationships is difficult and the probability of successful trail-and-error learning decreases when compared to routine, standardized tasks (Wood, 1986). Wood proposed that task complexity was a function of the extent to which a task requires a number of distinct behaviors, the number of choices required to perform the task, and the degree of uncertainty involved in performance of the task. On the basis of this reasoning, Heggestad and Kanfer (2005) posited that performance should not predict efficacy as strongly in complex tasks as it does in routine tasks. However, more accurate feedback regarding prior performance—such as that available during objective AARs—should

provide an opportunity to discover such cause-and-effect relationships regarding previous behaviors and performance outcomes.

Based on the aforementioned theories of self-regulation (Bandura, 1997), the role of the accuracy of feedback in the AAR process (Villado, 2008), and the nature of the performance-efficacy relationship (Lindsley et al., 1995) in the previously discussed sections, the following is hypothesized:

Hypothesis 1b: The team performance and team-efficacy relationship will be moderated by the level of objectivity of an AAR. Specifically, this relationship will be stronger for teams trained using objective AARs than for teams trained using subjective AARs, and stronger for teams trained using subjective AARs than for teams trained without AARs.

The Role of Time in the Performance-Efficacy Relationship

Time is acknowledged to play an important role in organizational research, particularly between predictors and performance criteria (Ancona, Goodman, Lawrence, & Tushman, 2001; Arthur et al., 2007; Kozlowski & Klein, 2000; Mitchell & James, 2001). As such, its impact on the relationship between performance and efficacy is of interest.

Mitchell and James (2001) posit that theory has tended to involve explanations and predictions of simple relationships between two variables (e.g., performance and accuracy) at a static point in time—a cross-sectional research design. However, as many scholars (e.g., Ancona, Okhuysen, & Perlow, 2001; George & Jones, 2000; Mitchell & James, 2001) have noted, time is a critical factor in explaining relationships between

variables of interest because relationships may change over time. For example, the relationship between performance and efficacy may change as a function of time, and furthermore, this relationship may change across time differently at different levels of a third variable (e.g., accuracy of performance feedback). In the performance-efficacy literature, studies at the individual (Feltz, 1988; Heggestad & Kanfer, 2005; Locke, Frederick, Lee, & Bobko, 1984; Shea & Howell, 2000) and the team levels (Arthur et al., 2007) have demonstrated that whereas the predictive validity of performance on efficacy remains relatively constant across trials, the predictive validity of efficacy on subsequent performance decreases across time.

Changes in the predictive validity of efficacy on performance across time may be explained by the process by which efficacy beliefs emerge. Bandura (1986) noted that efficacy beliefs change over time to reflect the most recent performance experience, and this view is supported by Gist and Mitchell (1992) who posit that efficacy is a dynamic construct that constantly changes as new information and experiences are acquired. Enactive mastery—a primary source of efficacy—is based on Guthrie's (1935) recency principle, which suggests that individuals confronted with a situation closely resembling an earlier situation are more likely to behave in the same manner as they did in the earlier situation. Because efficacy is task-specific (Bandura, 1997), efficacy is more likely to reflect a more recent performance experience than a more distal one.

Shea and Howell (2000) state that Bandura's (1986) idea that enactive attainment leads to mastery implies that the performance-efficacy relationship should have a general, uninterrupted upward or downward trend. Specifically, the reciprocal

relationship in which performance influences an efficacy judgment, which subsequently influences performance, can be interpreted as a performance-efficacy relationship that, over time, is characterized by monotonically increasing or decreasing levels of efficacy and performance—what Lindsley et al. (1995) refer to as an upward or downward spiral. Bandura's (1986) position that mastery is progressive and is derived from current performance outcomes also implies that judgments of efficacy are also progressive and are derived from current performance outcomes. This suggests that the relationship between performance and efficacy is influenced by time and each efficacy belief should be more strongly predicted by the most immediately preceding performance episode than by more distal preceding performance episodes. Studies at the individual level have indicated that efficacy is moderately correlated with prior performance at both the individual (Ackerman, Kanfer, & Goff, 1995, *r*s ranges from .38 to .43; Heggestad & Kanfer, 2005, *r*s ranged from .32 to .68) and team levels (Arthur et al., 2007, *r*s ranged from .39 to .62; Katz-Navon & Erez, 2005, $\beta = .56$). Because more accurate feedback should result in efficacy beliefs that more accurately reflect performance, and because of the progressive nature of efficacy beliefs, it is hypothesized that:

Hypothesis 2a: Team-efficacy will be better predicted by more proximal team performance than by more distal team performance.

In addition, more accurate feedback will result in team performance more strongly predicting team-efficacy, and the benefits of this more accurate feedback will accrue across time. Therefore, it is hypothesized:

Hypothesis 2b: The increase in the predictive validity of team performance on team-efficacy will be greater for teams trained using objective AARs than for teams trained using subjective AARs, and greater for teams trained using subjective AARs than for teams trained without AARs.

Within-Team Agreement in the Context of AARs

Team-efficacy is an aggregate measure comprised of individual team members' perceptions of their team's efficacy. As such, agreement of team members is of interest. Within-team agreement on a referent-shift consensus operationalization of team-efficacy improves over time because continued interaction among team members allows those team members to better estimate the capabilities of the team (Arthur et al., 2007; Baker, 2001; Jung & Sosik, 1999, 2003). During an AAR, team members interact with each other to review prior performance. During this interaction, team members not only receive feedback about their performance, they also gain information about other team members' capabilities and performance. Furthermore, because the AAR process provides a forum for information sharing (Mesmer-Magnus & DeChurch, 2009), team-efficacy judgments in the context of an AAR will be a function of this greater degree of shared information. In addition, team members rely on previous team performance and reflection on such performance to assess their team's efficacy. In both objective and subjective AARs, the availability of feedback and opportunity for information sharing and reflection provides a foundation for more accurate assessments of team-efficacy. Furthermore, the opportunity for information sharing afforded by AARs (regardless of

the level of objectivity) should increase within-team agreement about team-efficacy. Based on this reasoning, it is hypothesized that:

Hypothesis 3: Within-team agreement of team-efficacy ratings will increase over time and this increase will be stronger for teams that engage in an AAR, regardless of the level of objectivity of the AAR.

The Present Study

Performance and efficacy are reciprocally causal; however, the effect of performance on subsequent perceptions of efficacy has received little attention, especially in the context of team training. In addition, the moderating effect of feedback accuracy on the relationship between team performance and team-efficacy is largely unexplored. As such, the objective of the present study was to investigate the relationship between team performance and team-efficacy in the context of AARs. Specifically, this study examined the conjoint influence of (a) the accuracy of performance feedback available to trainees during AARs, and (b) time on the predictive validity of team performance on team-efficacy.

METHOD

Participants

Participants were recruited from the human subjects pool of Texas A&M University's psychology department. The sample consisted of 492 individuals (47.36% female) who participated in 123 mixed-sex 4-person teams. Participants reported a mean age of 18.84 yr (*SD* = 1.27) and described themselves as having average video-game experience (*M* = 1.81, *SD* = 0.65; video-game experience was measured using a 3-point scale where 1 = novice, 2 = average, and 3 = expert). Participants were provided with course credit for their participation. Additionally, to motivate them to remain focused and attempt to improve their performance during the study, participants in the first, second, and third highest performing teams in each of the three conditions were awarded $80, $40, and $20, respectively. Overall and condition-specific demographic information are presented in Table 1.

To assess the probability of detecting the effects indicated by the hypotheses, a power analysis was conducted using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) and Cohen's (1992) guidelines for small, medium, and large effect sizes (f^2 = .03, .15, and .35, respectively). This power analysis indicates that—with the current sample size of 123 teams and an alpha of .05—a test that regresses team-efficacy on team performance, AAR condition, and time results in a 32% chance of detecting a small effect. To achieve a power level of .80, a sample size of 368 teams would be necessary to detect the same size effect. However, the conditions of the present study (i.e., N = 123, α = .05) result in a 96% chance of detecting a medium effect.

Table 1

Demographic Composition of the Sample by Training Condition

	Training Condition							
	Non-AAR		Subjective AAR		Objective AAR		Overall	
	n	%	*n*	%	*n*	%	*n*	%
Sex								
Female	89	51.74	71	44.38	73	45.63	233	47.36
Male	83	48.26	89	55.63	87	54.38	259	52.64
Number of Males per Team								
0	3	6.98	2	5.00	2	5.00	7	5.69
1	13	30.23	6	15.00	9	22.50	28	22.76
2	15	34.88	16	40.00	14	35.00	45	39.59
3	8	18.60	13	32.50	10	25.00	31	25.20
4	4	9.30	3	7.50	5	12.50	12	9.76
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Age	18.72	0.87	18.93	1.72	18.87	1.09	18.84	1.27
Video-game experience	1.86	0.66	1.80	0.60	1.77	0.68	1.81	0.65

Note. *N* = 492. AAR = after-action review.

Thus, the study sample size of 123 teams provides sufficient power (power = .96) to detect a medium effect, but insufficient power (power = .32) to detect a small effect.

Measures

Performance task—Steel Beasts Pro PE. Steel Beasts Pro PE ver. 2.370 (eSim Games, 2007) was used to assess team performance. *Steel Beasts Pro PE* is a cognitively complex, PC-based tank synthetic task environment, allowing multiple players to be networked together to cooperatively complete a mission in a simulated battlefield environment. The simulator uses highly accurate replicas of U.S. M1A1 and Russian T-72 tanks to simulate an armored warfare environment (see Figure 2).

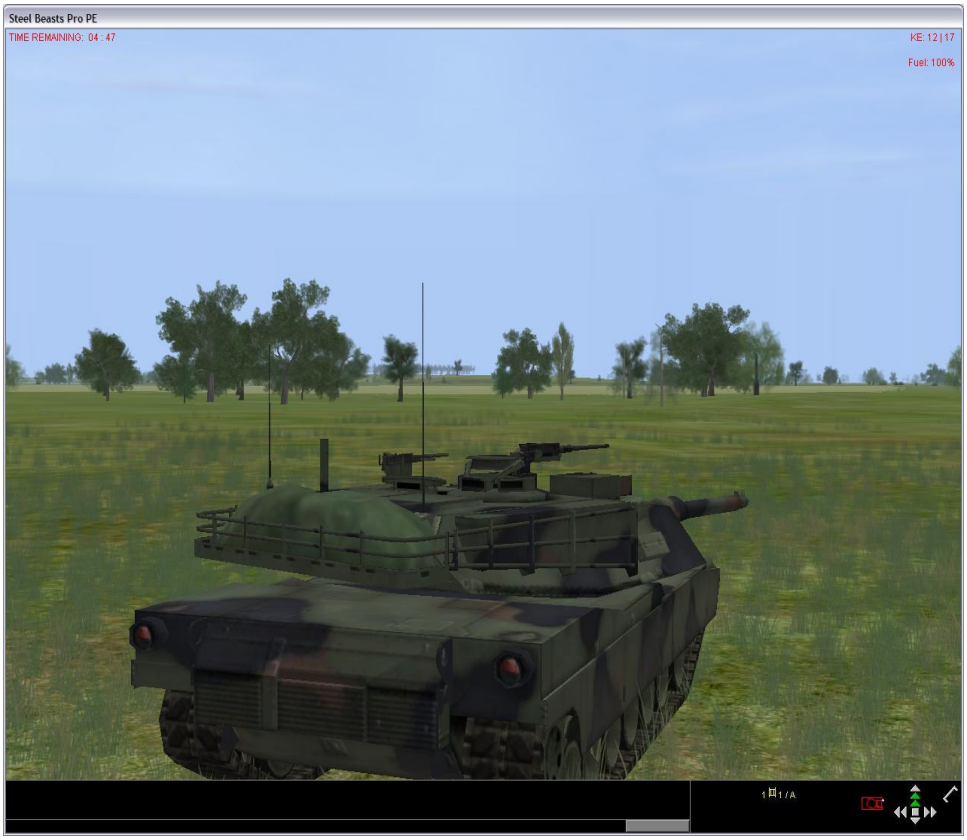


Figure 2. Steel Beasts replica M1A1 tank.

Participants operated the PC-based simulator using a monitor, keyboard, mouse, and joystick. The simulated environment consisted of a two-tank platoon of U.S. M1A1 tanks controlled by the participants. Four networked computers were used to operate the two-tank platoon; each participant had his/her own computer. Each tank in the platoon was operated by two participants; one participant served as the gunner and a second participant served as the commander/driver of the tank. Therefore, each team was comprised of two gunners and two commander/drivers (see Figure 3). Team members communicated with each other via voice-activated microphones and headphones.

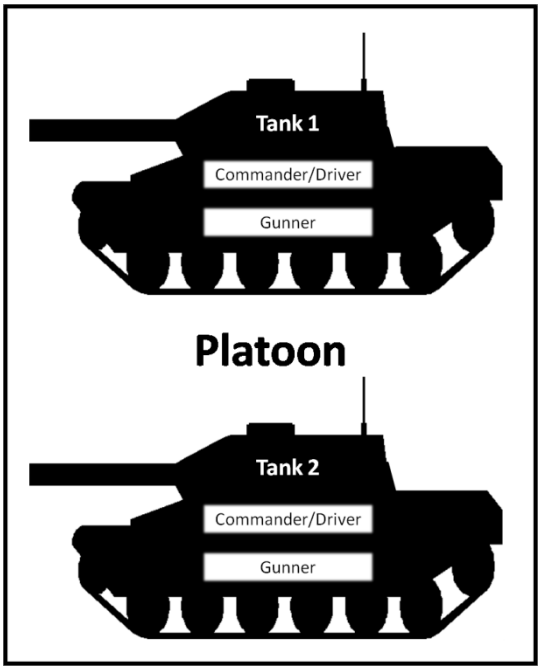


Figure 3. Illustration of the roles within and between tanks.

Multiple first-person perspective views were available to each participant, depending on his or her role. For example, gunners were able to switch between multiple gun sight views and a map view of the battlefield. Commander/drivers were able to switch between several views ranging from sitting inside the tank to standing up through the hatch of the tank, in addition to a view of the gunner's gun sight, and a map view of the battlefield.

The performance task was highly interdependent, with elements of both task and outcome interdependency. Task interdependency existed at the level of the tank such that the tank could not be operated successfully without the combined effort of the gunner and commander/driver. Outcome interdependency existed at the level of the

team. Specifically, missions were designed such that a single tank was not able to complete the mission without the assistance of the other tank.

Task interdependence was verified via ratings of team interdependency (Arthur, Edwards, Bell, Villado, & Bennett, 2005; Arthur et al., in press). Specifically, participants provided holistic ratings of (a) team-relatedness (i.e., the extent to which working with platoon members is required for optimal performance on Steel Beasts) and (b) team workflow (i.e., the manner in which work between platoon members flows for optimal performance on Steel Beasts). Holistic team-relatedness (*M* = 4.56, *SD* = 0.74; 1 = "Not required to work with platoon members at all for optimal performance", 5 = "Very much required to work with platoon members for optimal team performance) and team workflow ratings (*M* = 4.71, *SD* = 0.75; 1 = "Not a team job", 5 = "Intensive interdependence") indicated that Steel Beasts requires high levels of interdependence for optimal performance. A detailed description of the team-relatedness and team workflow scales is described in Arthur et al. (2005).

Steel Beasts Pro PE missions. There were two test missions for each of the three sessions. The same map was used for all six test missions (see Figure 4). Participants also completed two practice missions, one during Session 2 and the second during Session 3 (see Table 2). The practice missions were identical to the test missions; however, the participants were told that their scores on the practice missions would not count towards their performance scores. Each mission required a team to destroy 10 enemy tanks while the participants were en route to a target destination. Missions (both practice and test) consisted of a 2-minute briefing and planning session during which

teams were shown a mission briefing with information regarding mission objectives and rules, potential enemy positions, and enemy capabilities. Teams were then encouraged to formulate a strategy to complete the mission. After the briefing and planning session, for test missions, teams were allowed 10 min to complete the mission. A mission ended when (a) the team completed all mission objectives, (b) all participant tanks were destroyed, or (c) the 10-minute time limit expired. In contrast to the test missions, for practice missions, teams were allowed 15 min which they could use for either planning or interacting with the simulator. The first practice provided the participants with suggested waypoints for optimal performance of the missions (see Figure 5), whereas the second practice session was identical to the test missions with the exception of the time limit.

Performance scores were obtained at the team level. Teams earned points for the number of enemy tanks destroyed (5 points per tank) and for advancing beyond certain boundaries (2.5 points per tank per boundary crossed [e.g., each horizontal dashed line in Figure 4] and 12.5 for each tank that reached the objective). Teams lost points for destroying one of their own tanks (-50 points per fratricide). Thus, the total possible points ranged from -50 to 100. As previously noted, team performance for each session was operationalized as the average of the team's scores for the two test missions that were performed in each session. The method used to determine performance scores was explained to participants during each mission briefing and performance scores were available for them to review at the conclusion of every mission.

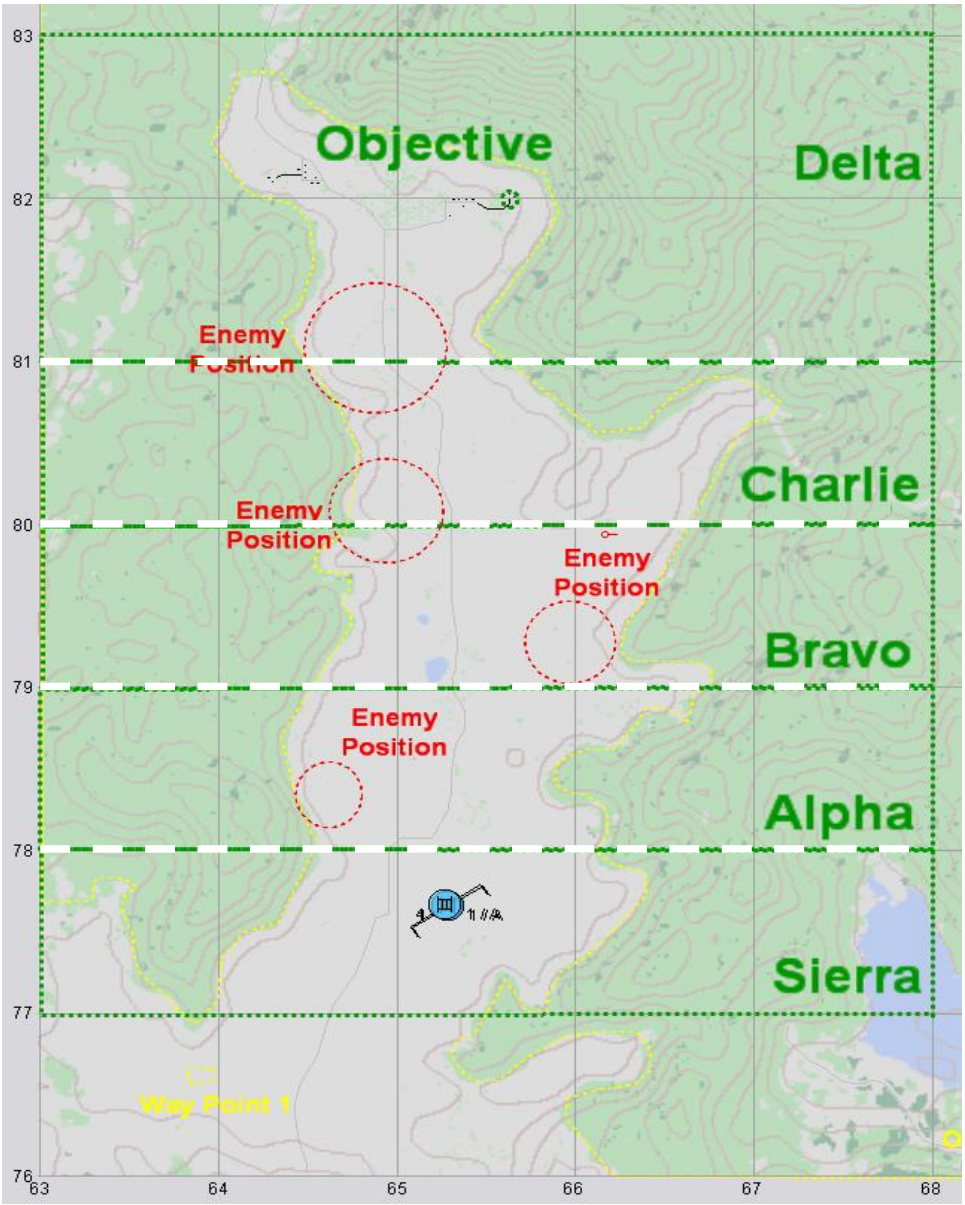


Figure 4. Mission map for test missions and the second practice mission.

Table 2

Schedule of Activities for Each Training Session by Training Condition

Session	Scheduled Activities		
0	Informed consent Team assigned to training condition Participants assigned to team Video-game experience Demographics Individual tutorials		
	Training Condition		
	Non-AAR (*n* = 43 teams)	**Subjective AAR (*n* = 40 teams)**	**Objective AAR (*n* = 40 teams)**
1	Planning Test Mission 1[a] Planning Test Mission 2[a] Filler Task Team-efficacy (Time 1)	Planning Test Mission 1[a] Planning Test Mission 2[a] AAR Team-efficacy (Time 1)	Planning Test Mission 1[a] Planning Test Mission 2[a] AAR Team-efficacy (Time 1)
2	Planning Practice Mission 1 Filler Task Planning Test Mission 3[b] Planning Test Mission 4[b] Filler Task	Planning Practice Mission 1 AAR Planning Test Mission 3[b] Planning Test Mission 4[b] AAR	Planning Practice Mission 1 AAR Planning Test Mission 3[b] Planning Test Mission 4[b] AAR
3	Planning Practice Mission 2 Filler Task Planning Test Mission 5[c] Planning Test Mission 6[c] Team-efficacy (Time 3)	Planning Practice Mission 2 AAR Planning Test Mission 5[c] Planning Test Mission 6[c] Team-efficacy (Time 3)	Planning Practice Mission 2 AAR Planning Test Mission 5[c] Planning Test Mission 6[c] Team-efficacy (Time 3)

Note. AAR = after-action review. Planning periods were limited to 2 min, test missions were limited to 10 min, practice missions were limited to 15 min, and AARs were limited to 10 min. Team-efficacy was not measured during Session 2 (Time 2). [a] A team's performance score at Time 1 is the mean of Test Missions 1 and 2. [b] A team's performance score at Time 2 is the mean of Test Missions 3 and 4. [c] A team's performance score at Time 3 is the mean of Test Missions 5 and 6.

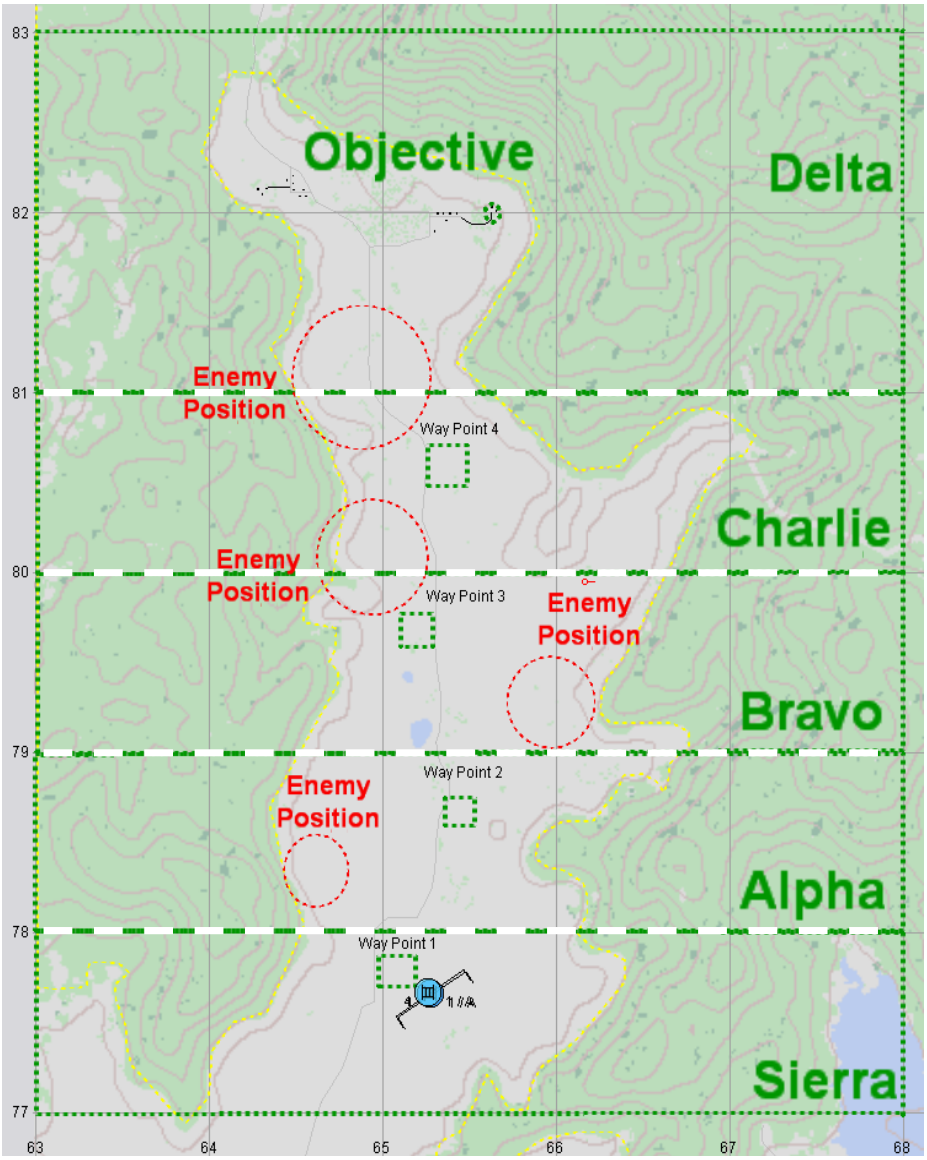


Figure 5. Mission map for first practice mission.

Team-efficacy. A modified version of the Arthur et al.'s (2007) team-efficacy measure was used to assess team-efficacy. The measure consisted of six task-specific items with a team referent. Participants provided their ratings using a 5-point rating scale (1 = strongly disagree, 5 = strongly agree). Team-efficacy scores were calculated using the mean of the mean individual-level item responses. Internal consistency estimates for the first and second administrations of the team-efficacy scores at the individual-level of analysis were .92, and .93, respectively (N = 492). These estimates are similar in magnitude to those reported by Arthur et al. (2007; mean coefficient alpha of .81) and Villado (2008; mean coefficient alpha of .93). The team-efficacy measure is presented in Appendix A.

Demographics. Participants reported their age, sex, experience with video games, and whether they had previous experience with Steel Beasts. A single video game experience item asked participants to describe their general experience with video games using a 3-point scale (i.e., novice, average, expert). Prior experience with Steel Beasts was collected with the intention of eliminating participants who had prior experience with the task. However, no participant reported any prior experience with Steel Beasts, and so no one was removed from the study for this reason. The demographics measure is presented in Appendix B and the video game experience measure is presented in Appendix C.

Design and Procedure

This study utilized a 3 (training condition: non-AAR, versus subjective AAR, versus objective AAR) × 3 (sessions) repeated measures design. Training condition

served as the between-subjects independent variable, and session served as the repeated or within-subjects independent variable. Two dependent variables (team performance and team-efficacy) were measured periodically throughout the study protocol to assess skill and efficacy in teams. An overview and summary of the experimental protocol is presented in Table 2.

The study protocol was 5 hr long and was divided into three phases. During the first phase, participants were familiarized with the protocol and completed the informed consent form and the demographics measure. After completing the measures, participants were then randomly assigned to a specific role within the team as either the gunner or commander/driver. The team was then randomly assigned to a training condition (i.e., non-AAR, subjective AAR, or objective AAR).

Training Manipulation

Participants were trained to operate the simulator first as individuals and then as a team. During the individual training phase, participants were allowed 45 min to complete 9 training tutorials. Each tutorial began with participants reading the tutorial content from a tutorial handbook. Once participants understood the content and objectives of the tutorial, they then completed a mission that provided hands-on practice of the tutorial content. Subsequent tutorials continued using the same procedure. Seven of the training tutorials focused on tasks relevant to a participant's role and the remaining two tutorials focused on tasks relevant to their teammate's role.

During the team training and performance phase, participants operated the simulator as a team to complete the six team-based missions. All participants completed

the same test missions in the same order, regardless of training condition. The events that followed each team performance mission depended on the training condition to which the team was assigned.

Non-AAR training condition. Once a two-mission session ended, participants assigned to the control training condition were given the opportunity to view their numerical performance score and then completed the specified paper-and-pencil measures as per Table 2. After completing the measures, participants did not participate in an AAR, but instead completed a filler task that was unrelated to Steel Beasts. This was to ensure similar spacing between missions for the non-AAR and AAR training conditions.

Subjective AAR training condition. After completing a two-mission session, participants in the subjective AAR training condition were given the opportunity to view their numerical performance score and then participated in a 10-minute AAR, monitored by one of two facilitators. Prior to the first AAR, the facilitator explained the AAR process to team members and provided teams with a form (presented in Appendix D) that detailed each step of the AAR process. After introducing participants to the AAR process, facilitators only intervened during AARs to ensure that teams completed each step of the AAR in the order presented in Figure 1 within the specified time limits. Thus, the AAR facilitator did not assume an instructor role during the AARs; teams participated in each AAR using a self-managed approach.

Subjective AARs began with participants recalling the intended outcome and the actual outcome of their most recently completed mission. Participants then compared

the two to determine whether their goals had been met. Next, participants identified specific behaviors or events that contributed to or detracted from achieving the mission objectives. The participants were then encouraged to set specific and difficult, yet attainable goals for the subsequent mission. Each AAR concluded with participants identifying behaviors and actions that would increase the likelihood of meeting their self-set goals and subsequent mission objectives. Teams then completed the specified paper-and-pencil measures as per Table 2.

Objective AAR training condition. Participants assigned to the objective AAR training condition were given the opportunity to view their numerical performance score and then participated in an AAR after each two-mission session in the same manner (and within the same 10-minute time period) as participants in the subjective AAR training condition. The AAR form is presented in Appendix D. However, participants in the objective AAR training condition objectively reviewed the progress of their most recently completed mission using the simulator's review tool, operated by the facilitator. The review tool allowed participants to replay, pause, and move forward or backward through the simulated environment of the most recently completed mission. Participants could view the mission progress from multiple perspectives and examine it from any point in the simulated environment (e.g., from either tanks' perspective, the enemy's perspective, or a top-down view of the mission). After the AAR, teams completed the specified paper-and-pencil measures as per Table 2.

RESULTS

Prior to analyzing data at the team level, individual-level data (i.e., team-efficacy ratings) were evaluated to justify aggregation to the team level. Agreement and reliability indices (i.e., ICC(1), ICC(2), and $r_{wg(j)}$) were calculated to assess the appropriateness of aggregating individual-level data to the team-level (Bartko, 1976; James, Demaree, & Wolf, 1984). Reliability estimates and mean and median agreement indices, which are presented in Table 3, indicated that aggregation to the team level was justified. Therefore, team-level ratings of team-efficacy were created by averaging individuals' team-efficacy ratings within teams. Team performance scores were recorded at the team level and thus did not require aggregation.

Table 3

Team-Efficacy Intraclass Correlation Coefficients and Mean Agreement Indices

			$r_{wg(j)}$	
Administration	ICC(1)	ICC(2)	Mean	Median
Time 1	.11	.49	.86	.91
Time 3	.14	.63	.83	.90

Note. N = 492. k = 123 teams. ICC(1) and ICC(2) were calculated using formulas presented by Bartko (1976). $r_{wg(j)}$ was calculated using the formula presented by James et al. (1984).

In addition, these data were collected as part of a larger data collection effort in which the comparative effectiveness of AARs in co-located and distributed training environments was investigated. As such, teams were randomly assigned to an AAR condition (i.e., non-, subjective, and objective AAR) and were also randomly assigned to

either a geographically co-located or distributed training environment. Because the geographic distribution of team members was not of substantive interest in this study, teams were not differentiated on the basis the geographic distribution of team members. However, before collapsing geographically co-located and distributed teams into a single group, a series of analyses of covariance (ANCOVAs) were conducted to determine whether the interaction between AAR condition and geographic training location significantly predicted any variance in the variables of interest in this study (i.e., team performance, team-efficacy, or within-team agreement about team-efficacy). The results of this series of analyses are presented in Appendix E and indicate that none of the variables of interest in the current study differed on the basis of the interaction between AAR condition and geographic training location. In addition, individuals were randomly assigned to teams, and teams were randomly assigned to both AAR condition and geographic training location—all factors serving to mitigate the presence of preexisting group differences (Miller & Chapman, 2001). On the basis of the aforementioned evidence and arguments, teams in co-located and distributed training environments were collapsed within each AAR condition.

Descriptive Statistics

The means, standard deviations, and correlations among the study variables collapsed across AAR conditions are presented in Table 4. The means, standard deviations, and correlations among the study variables by AAR condition are presented in Table 5. Both team performance and team-efficacy increased across time. Analysis of the correlation matrix indicates a simplex pattern between the team performance

scores such that the largest correlations exist between temporally adjacent performance scores and correlations decrease in magnitude as the number of intervening performance sessions increases (Arthur et al., 2007). In addition, the correlations between performance and team-efficacy display a similar pattern. Specifically, team-efficacy at Time 1 is more strongly related to team performance at Time 2 (r = .24) than team performance at Time 3 (r = .09). Similarly, team-efficacy at Time 3 is most strongly correlated with team performance at Time 3 (r = .52), less strongly correlated with team performance at Time 2 (r = .20), and not related to team performance at Time 1 (r = -.01). In addition, the results indicate that overall, team-efficacy at the end of training (i.e., team-efficacy at Time 3) was correlated with both team performance at the end of training (i.e., team performance at Time 3, r = .52) and initial team-efficacy (i.e., team-efficacy at Time 1, r = .54).

Table 4

Descriptive Statistics and Correlations Between Study Variables Collapsed Across all AAR Conditions

Variable	*M*	*SD*	Perf 1	Eff 1	$r_{wg(j)}$ 1	Perf 2	Perf 3	Eff 3
Perf 1	27.93	6.33	—					
Eff 1	3.32	0.47	.13	.92				
$r_{wg(j)}$ 1	0.86	0.18	.03	.16*	—			
Perf 2	35.62	9.11	.24**	.25**	.15*	—		
Perf 3	39.33	10.57	.09	.15	-.02	.32**	—	
Eff 3	3.54	0.52	-.01	.54**	.03	.20*	.52**	.93
$r_{wg(j)}$ 3	0.83	0.22	.11	.08	.24**	.06	.20*	.21*

Note. $N = 123$. AAR = After-action review. Perf 1 = performance at Time 1 (i.e., the mean of team performance scores on Test Missions 1 and 2); Eff 1 = team-efficacy at Time 1 (i.e., after Test Mission 2); $r_{wg(j)}$ 1 = within-team agreement on team-efficacy at Time 1 (i.e., after Test Mission 2); Perf 2 = performance at Time 2 (i.e., the mean of team performance scores on Test Missions 3 and 4; Perf 3 = performance at Time 3 (i.e., the mean of team performance scores on Test Missions 5 and 6; Eff 3 = team-efficacy at Time 3 (i.e., after Test Mission 6); $r_{wg(j)}$ 3 = within-team agreement on team-efficacy at Time 3 (i.e., after Test Mission 6). The range of potential performance scores was from -50 to 100. Coefficient alpha reliabilities are located on the diagonal. * $p < .05$, ** $p < .01$, one-tailed.

Table 5

Correlations Between Study Variables by AAR Condition

	Non-AAR (N = 43)						Subjective AAR (N = 40)						Objective AAR (N = 40)					
Variable	Perf 1	Eff 1	$r_{wg(j)}$ 1	Perf 2	Perf 3	Eff 3	Perf 1	Eff 1	$r_{wg(j)}$ 1	Perf 2	Perf 3	Eff 3	Perf 1	Eff 1	$r_{wg(j)}$ 1	Perf 2	Perf 3	Eff 3
Perf 1	—						—						—					
Eff 1	.02	.92					.39**	.93					.03	.91				
$r_{wg(j)}$ 1	-.00	.23	—				.09	.54**	—				.32*	.13	—			
Perf 2	.26*	.15	-.08	—			.41**	.28*	.56**	—			.10	.28*	.11	—		
Perf 3	.29*	.18	.09	.45**	—		.25	.14	.16	.38**	—		-.19	-.15	-.20	.09	—	
Eff 3	.18	.51**	-.01	.28*	.56**	.92	.13	.52**	.36*	.29*	.32*	.95	-.19	.31*	-.10	-.16	.44**	.92
$r_{wg(j)}$ 3	.09	.03	.37**	-.02	.29*	.23	.24	.10	.18	.27*	.13	.35*	-.12	-.22	.17	-.26	.38**	.43**

Note. AAR = After-action review; Perf 1 = performance at Time 1 (i.e., the mean of team performance scores on Test Missions 1 and 2); Eff 1 = team-efficacy at Time 1 (i.e., after Test Mission 2); $r_{wg(j)}$ 1 = within-team agreement on team-efficacy at Time 1 (i.e., after Test Mission 2); Perf 2 = performance at Time 2 (i.e., the mean of team performance scores on Test Missions 3 and 4; Perf 3 = performance at Time 3 (i.e., the mean of team performance scores on Test Missions 5 and 6; Eff 3 = team-efficacy at Time 3 (i.e., after Test Mission 6); $r_{wg(j)}$ 3 = within-team agreement on team-efficacy at Time 3 (i.e., after Test Mission 6). The range of potential performance scores was from -50 to 100. Coefficient alpha reliabilities are located on the diagonals. * $p < .05$, ** $p < .01$, one-tailed.

Table 6

Means, Standard Deviations, and Effect Sizes for Study Variables by AAR Condition

	Non-AAR (N = 43)		Subjective AAR (N = 40)		Objective AAR (N = 40)		Effect size (d)		
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	Objective AAR vs Subjective AAR	Subjective AAR vs Non-AAR	Objective AAR vs Non-AAR
Perf 1	28.72	5.77	26.81	6.91	28.19	6.31	0.21	-0.30	-0.09
Eff 1	3.15	0.39	3.47	0.47	3.34	0.49	-0.26	0.75*	0.44
$r_{wg(j)}$ 1	0.79	0.25	0.89	0.14	0.91	0.06	0.26	0.50*	0.79*
Perf 2	33.84	9.37	35.84	8.94	37.31	8.88	0.16	0.22	0.38
Perf 3	36.10	9.94	40.97	9.81	41.16	11.36	0.02	0.49*	0.48*
Eff 3	3.30	0.54	3.72	0.39	3.63	0.51	-0.18	0.89*	0.65*
$r_{wg(j)}$ 3	0.79	0.24	0.82	0.23	0.88	0.16	0.27	0.16	0.46*

Note. AAR = After-action review; Perf 1 = performance at Time 1 (i.e., the mean of team performance scores on Test Missions 1 and 2); Eff 1 = team-efficacy at Time 1 (i.e., after Test Mission 2); $r_{wg(j)}$ 1 = within-team agreement on team-efficacy at Time 1 (i.e., after Test Mission 2); Perf 2 = performance at Time 2 (i.e., the mean of team performance scores on Test Missions 3 and 4; Perf 3 = performance at Time 3 (i.e., the mean of team performance scores on Test Missions 5 and 6; Eff 3 = team-efficacy at Time 3 (i.e., after Test Mission 6); $r_{wg(j)}$ 3 = within-team agreement on team-efficacy at Time 3 (i.e., after Test Mission 6). The range of potential performance scores was from -50 to 100. Effect sizes (*d*s) were computed by subtracting the second condition from the first such that a positive *d* indicates that participants in the first condition had higher scores than those in the second condition. * $p < .05$, ** $p < .01$, one-tailed.

The means, standard deviations, and effect sizes (*d*s) for the non-, subjective, and objective AAR conditions are presented in Table 6. Although there were no a priori hypotheses regarding differences between these conditions, several interesting findings are evident. There were no significant differences in team performance during the first two sessions. That is, the non-, subjective, and objective AAR conditions displayed similar levels of performance during each of the first two team performance sessions, but teams in the subjective AAR condition displayed higher levels of team-efficacy than teams in the non-AAR condition immediately following the initial performance session ($d = 0.75$). At the completion of training, both subjective and objective AAR conditions displayed higher performance levels than the non-AAR condition (*d*s of 0.49 and 0.48, respectively). In addition, at the completion of training, both subjective and objective AAR conditions displayed higher team-efficacy levels than the non-AAR condition (*d*s of 0.89 and 0.65, respectively). Although the subjective AAR condition displayed slightly higher team-efficacy than the objective AAR condition, this difference was not statistically significant.

Hypothesis Tests

Hypotheses 1a, 1b, 2a, and 2b involved testing the relationship between team performance and team-efficacy. These hypotheses were tested using partial least squares (PLS) analysis (Wold, 1985). PLS, in contrast to more well-known covariance-based structural equation modeling (SEM) techniques such as LISREL (Jöreskog & Sörbom, 1993) and Mplus (Muthen & Muthen, 1998), is a variance-based SEM technique used to analyze theoretically-derived models linking constructs (Fornell, 1982). Although

covariance- and variance-based SEM techniques are similar, each is used in support of different analysis objectives. For example, covariance-based SEM techniques are concerned with fitting covariance matrices and are more appropriate for fitting models based on strong theoretical groundings, whereas variance-based SEM techniques are concerned with prediction in a manner similar to ordinary least-squares regression (i.e., minimizing residual variances) and are appropriate for predictive research models tested in the early stages of theory development (Fornell & Bookstein, 1982). Fornell and Bookstein also note that covariance-based SEM techniques require multivariate normal data, relatively large sample sizes, and theoretically strong measurement and structural models. In contrast, variance-based SEM techniques such as PLS do not require multivariate normality, observation independence, or interval-level data (Faulk & Miller, 1992). In addition, PLS is suitable for use with small sample sizes (Barclay, Higgins, & Thompson, 1995). The path coefficients in a PLS structural model are standardized regression coefficients and are interpreted as such. To conduct significance tests of path coefficients within a single model and to conduct multiple-group comparisons, bootstrapping was used to generate *t* values and standard errors (Rigdon, Ringle, & Sarstedt, 2010). Unless otherwise specified, the standard errors of path coefficients were estimated using bootstrapping with a sample size of 1000. In order to compare the results of the variance-based PLS approach conducted using SmartPLS (Ringle, Wende, & Will, 2005) with a covariance-based SEM approach, similar analyses were conducted for Hypotheses 1a, 1b, 2a, and 2b using Mplus (Muthen & Muthen, 1998); the results of

the covariance-based SEM analyses were similar to those obtained from the variance-based PLS approach and are included in Appendix F for comparison purposes.

Hypotheses 1a and 1b. Hypothesis 1a stated that team performance would positively predict team-efficacy. A causal model depicting this relationship is presented in Figure 6. The model controlled for the strength of team-efficacy agreement by linking the strength of team-efficacy agreement at Time 1 with each measure of team performance and team-efficacy. Overall, although team performance at Time 1 positively predicted team-efficacy at Time 1, this relationship was not statistically significant, $\beta = .13$, $t(122) = 1.26$, $p = .10$. However; team performance at Time 3 did positively predict team-efficacy at Time 3, $\beta = .48$, $t(122) = 6.65$, $p < .05$. Thus, Hypothesis 1a was partially supported.

Hypothesis 1b stated that the relationship between team performance and team-efficacy would be moderated by the level of AAR objectivity. In particular, it was hypothesized that team performance would predict team-efficacy more strongly in the objective AAR condition than the subjective AAR condition and more strongly in the subjective AAR condition than the non-AAR condition. The path coefficients between team performance and team-efficacy in each AAR condition are shown in Figure 7.

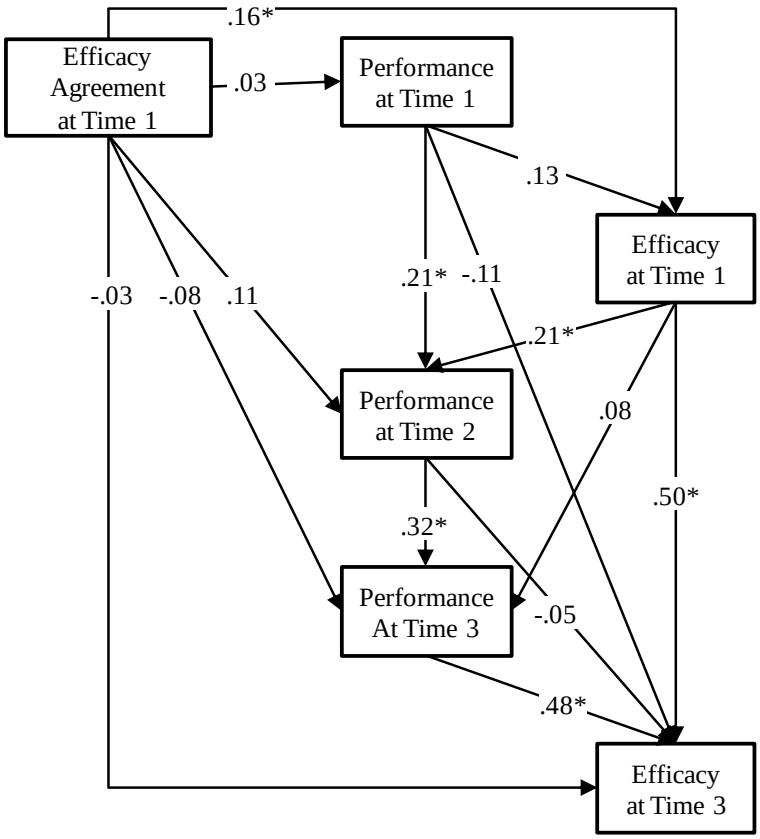


Figure 6. The relationship between team performance and team-efficacy over time. *$p < .05$.

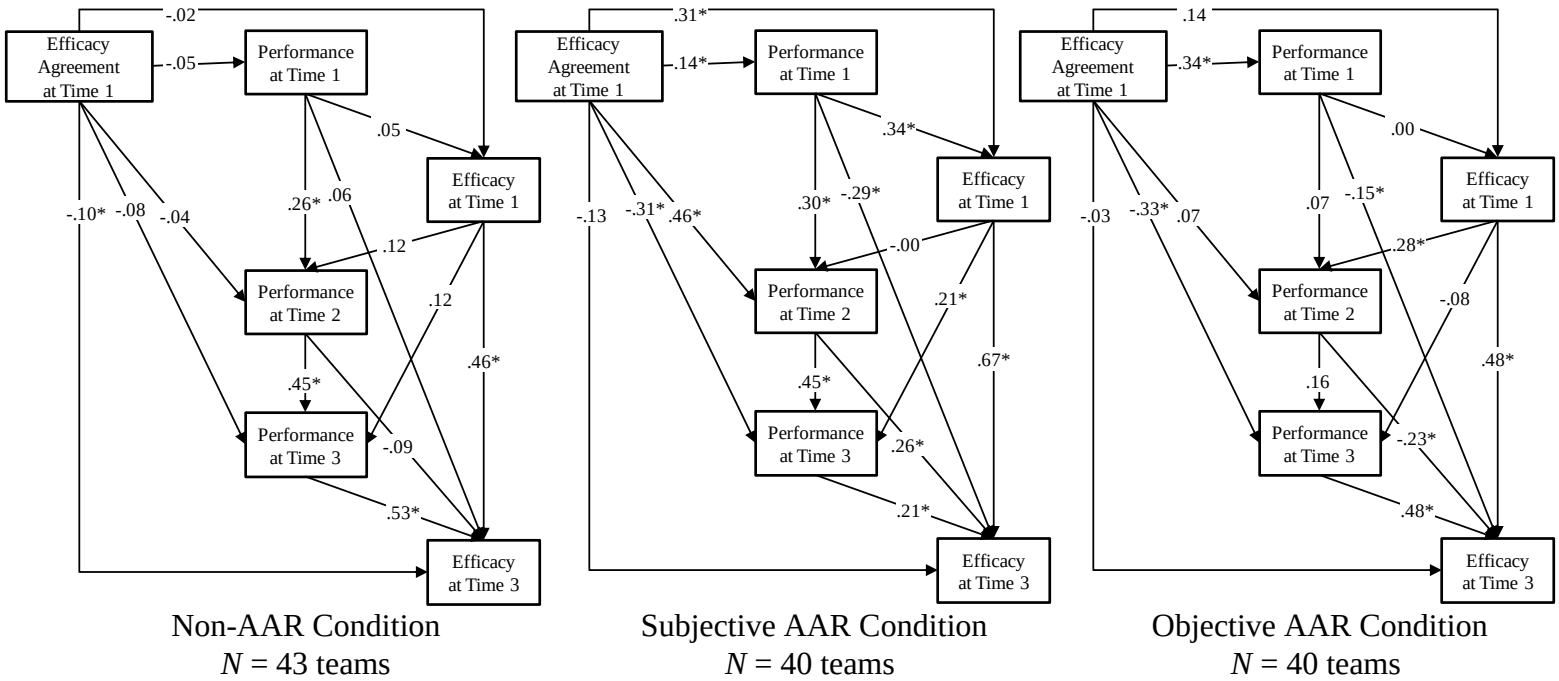


Figure 7. The relationship between team performance and team-efficacy over time for each after-action review (AAR) condition. *$p < .05$.

The moderating effect of AAR condition on teams' performance-efficacy relationships was examined using subsample analysis. Subsample analysis was conducted by (a) dividing the original sample into subgroups on the basis of a moderating variable (i.e., AAR condition), (b) conducting parallel model analysis for each level of the moderating variable, and (c) comparing the relationships found in each subgroup (see Sosik, Kahai, & Piovoso, 2009, for a primer on subsample analysis using PLS techniques in group and organizational research). Using PLS analysis, path coefficients and standard errors were estimated for the three AAR conditions and unpaired *t* tests were used to compare path coefficients according to the following formula (Chin, 2000):

$$t = \frac{b_1 - b_2}{\sqrt{\frac{(n_1 - 1)^2}{(n_1 + n_2 - 2)} * se(b_1)^2 + \frac{(n_2 - 1)^2}{(n_1 + n_2 - 2)} * se(b_2)^2} * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (1)$$

where b_1 and b_2 denote the parameter estimates of the path coefficients in subsample 1 and 2, n_1 and n_2 denote the number of observations in subsample 1 and 2, and $se(b_1)$ and $se(b_2)$ denote the standard errors of the path coefficients resulting from the bootstrapping procedure. The resultant test statistic is distributed as a *t* statistic with $(n_1 + n_2 - 2)$ degrees of freedom. The *t* values for comparisons of path coefficients between AAR conditions are presented in Table 7.

Contrary to Hypothesis 1b, which stated that team-efficacy would be predicted by team performance most strongly in the objective AAR condition and least strongly in the non-AAR condition, team performance did not predict team-efficacy as predicted. As seen in Table 7, team performance at Time 1 predicted team-efficacy at Time 1 more strongly in the subjective AAR condition than the non-AAR condition, $t(81) = 2.50$, $p < .01$, as expected. However, it predicted team-efficacy at Time 1 more weakly in the objective AAR condition that in both the subjective AAR and non-AAR conditions. Thus, at an early stage of training, performance predicted efficacy in the following manner: subjective AAR > non-AAR > objective AAR, as seen in Figure 7.

As predicted, team performance at Time 3 predicted team-efficacy at Time 3 more strongly in the objective AAR condition than in the subjective AAR condition, $t(78) = 2.50$, $p < .01$; however, it did not predict team-efficacy at Time 3 more strongly in the subjective AAR condition than in the non-AAR condition. Surprisingly, team-performance at Time 3 predicted team-efficacy at Time 3 most strongly in the non-AAR condition. Thus, at the later stage of training, performance predicted efficacy in the following manner: non-AAR > objective AAR > subjective AAR, as seen in Figure 7.

Taken together, the pattern of results observed in both early and later stages of training provides only weak support for Hypothesis 1b. In Table 7, two rows involving predictions of team-efficacy—rows 1 and 4—provide a summary of evidence relating to Hypothesis 1b. Of the six possible comparisons (i.e., objective AAR > subjective AAR, subjective AAR > non-AAR, and the implied objective AAR > non-AAR for each of the

two temporal relationships between performance and efficacy), only two relationships—indicated by positive *t* values—were in the hypothesized direction.

Table 7

Comparison of Path Coefficients Between AAR Conditions

Path			Comparison *t*-values		
			Objective AAR vs Subjective AAR	Subjective AAR vs Non-AAR	Objective AAR vs Non-AAR
1. Team Performance at Time 1	to	Team-Efficacy at Time 1	-2.67 (*df* = 78)	2.50** (*df* = 81)	-0.36 (*df* = 81)
2. Team Performance at Time 1	to	Team-Efficacy at Time 3	-1.00 (*df* = 78)	1.82* (*df* = 81)	0.91 (*df* = 81)
3. Team Performance at Time 2	to	Team-Efficacy at Time 3	-0.21 (*df* = 78)	1.30 (*df* = 81)	1.23 (*df* = 81)
4. Team Performance at Time 3	to	Team-Efficacy at Time 3	2.50** (*df* = 78)	-2.77 (*df* = 81)	-0.50 (*df* = 81)

Note. AAR = After-action review. Comparison statistics were computed by subtracting the second condition from the first such that a positive *t* statistic indicates the participants in the first condition had higher scores those in the second condition (i.e., the results were in the hypothesized direction).
* $p < .05$, ** $p < .01$, one-tailed.

Hypotheses 2a and 2b. Hypothesis 2a stated that team-efficacy would be predicted more strongly by more proximal team performance than by more distal team performance. Path coefficients indicating the relationships of interest can be seen in Figure 6. As seen in Figure 6, team-efficacy at Time 3 was positively predicted by team performance at Time 3, $\beta = .48$, $t(122) = 6.34$, $p < .05$, was predicted less strongly by team performance at Time 2, $\beta = .10$, $t(122) = 1.00$, $p = .16$, and was predicted most

weakly by team performance at Time 1, $\beta = -.02$, $t(122) = 0.19$, $p = .42$. Thus, team-efficacy at Time 3 was predicted most strongly by the most proximal performance score (Time 3) and most weakly by the most distal performance score (Time 1), which provided support for Hypothesis 2a.

Hypothesis 2b stated that the predictive validity of the most proximal team performance score on team-efficacy should be stronger in the objective AAR condition than in the subjective AAR condition and stronger in the subjective AAR condition than in the non-AAR condition. Furthermore, it also predicted the same pattern of predictive validity should hold for each successively more distal team performance score. Support for this hypothesis was mixed.

Regarding the most proximal performance episode (i.e., team performance at Time 3), team-efficacy at Time 3 was predicted more strongly by team performance in the objective AAR condition than in the subjective AAR condition, $t(78) = 2.50$, $p < .01$, as expected. However, it was not predicted more strongly by team performance in the subjective AAR condition than in the non-AAR condition. Surprisingly, team performance in the non-AAR condition predicted team-efficacy even more strongly than it did in the objective AAR condition. In summary, for the most proximal performance episode, the predictive validity of team performance on team-efficacy was as follows: non-AAR > objective AAR > subjective AAR, as seen in Figure 7.

Regarding the next more distal performance episode (i.e., team performance at Time 2), team-efficacy at Time 3 was not predicted more strongly by team performance in the objective AAR condition than in the subjective AAR condition. The magnitude of

the relationship between performance and efficacy was similar in the objective and subjective AAR conditions, but quite surprisingly this relationship in the objective AAR condition was negative, indicating that teams with lower performance at Time 2 tended to have higher efficacy scores at the end of training (i.e., Time 3). Finally, team-efficacy at Time 3 was predicted more strongly by team performance in the subjective AAR condition than in the non-AAR condition; however, this observed difference in predictive validities was not statistically significant. In summary, for the next more distal performance episode, the predictive validity of team performance on team-efficacy was as follows: subjective AAR > objective AAR > non-AAR, as seen in Figure 7.

Finally, regarding the most distal performance episode (i.e., team performance at Time 1), team-efficacy at Time 3 was not predicted more strongly by team performance in the objective AAR condition than in the subjective AAR condition. As expected, team-efficacy at Time 3 was predicted more strongly by team performance in the subjective AAR condition than in the non-AAR condition, $t(81) = 1.82$, $p < .05$. However, this relationship in the subjective AAR condition was negative, indicating that—in the subjective AAR condition—teams with lower performance scores early in training ended training with higher efficacy scores. In summary, for the most distal performance episode, the predictive validity of team performance on team-efficacy was as follows: subjective AAR > objective AAR > non-AAR, as seen in Figure 7.

Taken together, the previously discussed results indicate only partial support for Hypothesis 2b. In Table 7, the three rows involving predictions of team-efficacy at Time 3 (i.e., rows 2, 3, and 4) provide a summary of evidence relating to Hypothesis 2b.

Of the nine possible comparisons (i.e., objective AAR > subjective AAR, subjective AAR > non-AAR, and the implied objective AAR > non-AAR for each of the three temporal relationships between performance and efficacy), only five relationships—indicated by positive *t* values—were in the hypothesized direction; moreover, only two of these relationships were statistically significant (i.e., *t* values of 1.82 and 2.50).

Hypothesis 3. Hypothesis 3 stated that within-team agreement of team-efficacy ratings would increase over time and this increase would be stronger for teams that engaged in AARs, regardless of the level of objectivity of the AAR. In order to test this hypothesis, indices of within-team agreement (i.e., $r_{wg(j)}$) regarding team-efficacy were computed for each team at Time 1 and Time 3 and analyzed using a 2 × 2 mixed analysis of covariance (ANCOVA). AAR condition (i.e., non-AAR versus AAR) served as the between-subjects independent variable and time (Time 1 and Time 3) served as the within-subjects variable. Initial team performance was controlled by including it in the model as a covariate. This approach is advocated by Miller and Chapman (2001) as an appropriate analytical technique when individuals are randomly assigned to treatment groups and these groups do not exhibit differences on the covariate—conditions characteristic of the current study. The means and standard deviations of the indices of within-team agreement by time (Time 1 and Time 3) and AAR condition are presented in Table 8 and a graphical depiction of within-team agreement on team-efficacy across time is shown in Figure 8. The results presented in Table 8 indicate that from Time 1 to Time 3, within-team agreement on team-efficacy actually decreased for teams trained using AARs, whereas it remained constant for teams trained without AARs. However,

the results of the ANCOVA revealed that this decrease in within-team agreement did not differ across AAR conditions, $F(1, 120) = 0.86$, $p = .36$, $\eta^2 = .01$. Moreover, the within-subjects effect of time was not statistically significant, indicating that the observed overall decrease in within-team agreement on team-efficacy was not statistically significant, $F(1, 120) = 0.94$, $p = .33$, $\eta^2 = .01$. Thus, within-team agreement did not increase over time and Hypothesis 3 was not supported.

Table 8

Team-Efficacy Agreement by Time and AAR condition

	Condition						
	Non-AAR (N = 43)		AAR (N = 80)		Overall (N = 123)		
Administration	*M*	*SD*	*M*	*SD*	*M*	*SD*	*d*
Time 1	.79	0.25	.90	0.11	.86	0.18	0.71
Time 3	.79	0.24	.85	0.20	.83	0.22	0.25

Note. AAR = After-Action Review. Effect sizes (*d*s) were computed by subtracting the Non-AAR condition from the AAR condition such that a positive *d* indicates that participants in the AAR condition had higher scores than those in the Non-AAR condition.

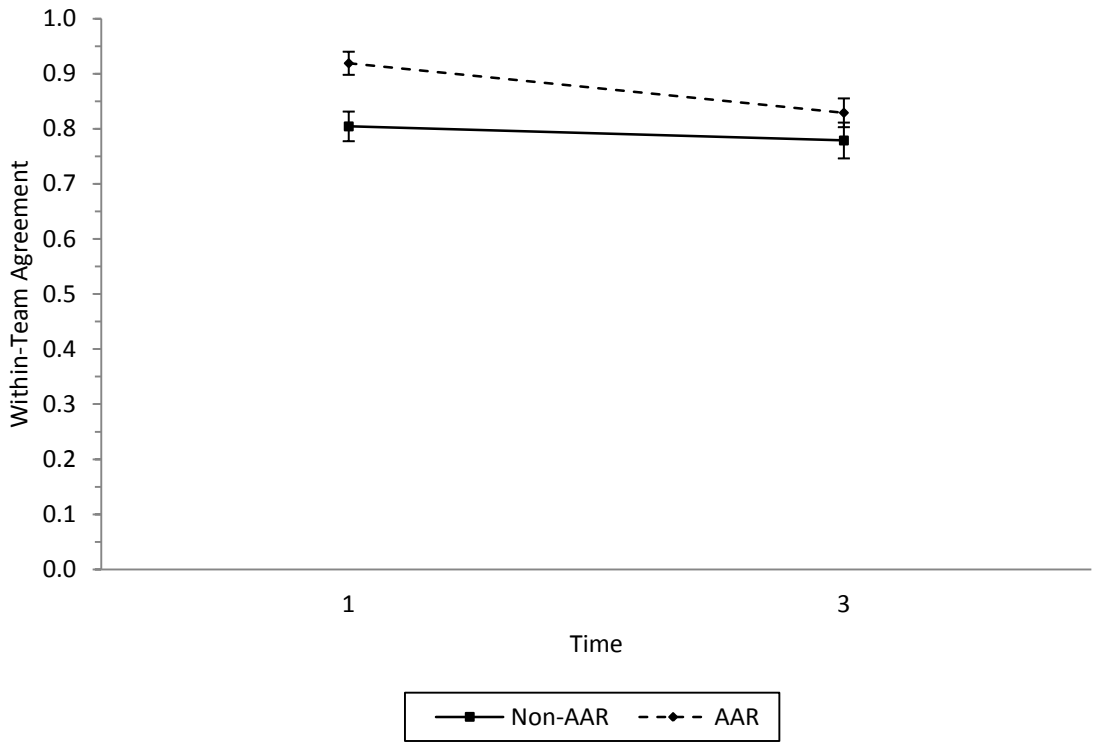


Figure 8. Within-team agreement on team-efficacy over time.

In addition, to further explore the pattern of results observed in within-team agreement among teams in the AAR condition, the AAR condition was disaggregated into subjective and objective AAR conditions and analyzed in a manner similar to the analysis used to test Hypothesis 3 (the single exception being that this analysis used a 3 × 2 mixed ANCOVA with AAR condition [i.e., non-AAR versus subjective AAR versus objective AAR] serving as the between-subjects independent variable).

The means and standard deviations of the indices of within-team agreement by time (Time 1 and Time 3) and AAR condition are presented in Table 9 and a graphical depiction of within-team agreement on team-efficacy across time is shown in Figure 9.

An ANCOVA based on the three original AAR conditions revealed neither a significant main effect of time nor a significant time × AAR condition interaction. However, the pattern of results seen in Table 9 and Figure 9 (i.e., the largest decrease in within-team agreement was observed in the subjective AAR condition) are tentatively suggestive of differences in the psychological mechanisms influencing agreement between the subjective and objective AAR conditions.

Table 9

Team-Efficacy Agreement by Time and Disaggregated AAR Condition

	Condition						*d*s		
	Non-AAR (*N* = 43)		Subjective AAR (N = 40)		Objective AAR (*N* = 40)		SUBJ vs NON	OBJ vs NON	OBJ vs SUBJ
Administration	*M*	*SD*	*M*	*SD*	*M*	*SD*			
Time 1	.79	0.25	.90	0.11	.86	0.18	0.60	0.32	-0.28
Time 3	.79	0.24	.85	0.20	.83	0.22	0.27	0.17	-0.10

Note. AAR = After-Action Review. NON = Non-AAR condition. SUBJ = Subjective AAR condition. OBJ = Objective AAR condition. Effect sizes (*d*s) were computed by subtracting the second condition from the first such that a positive *d* indicates that participants in the first condition had higher scores than those in the second condition.

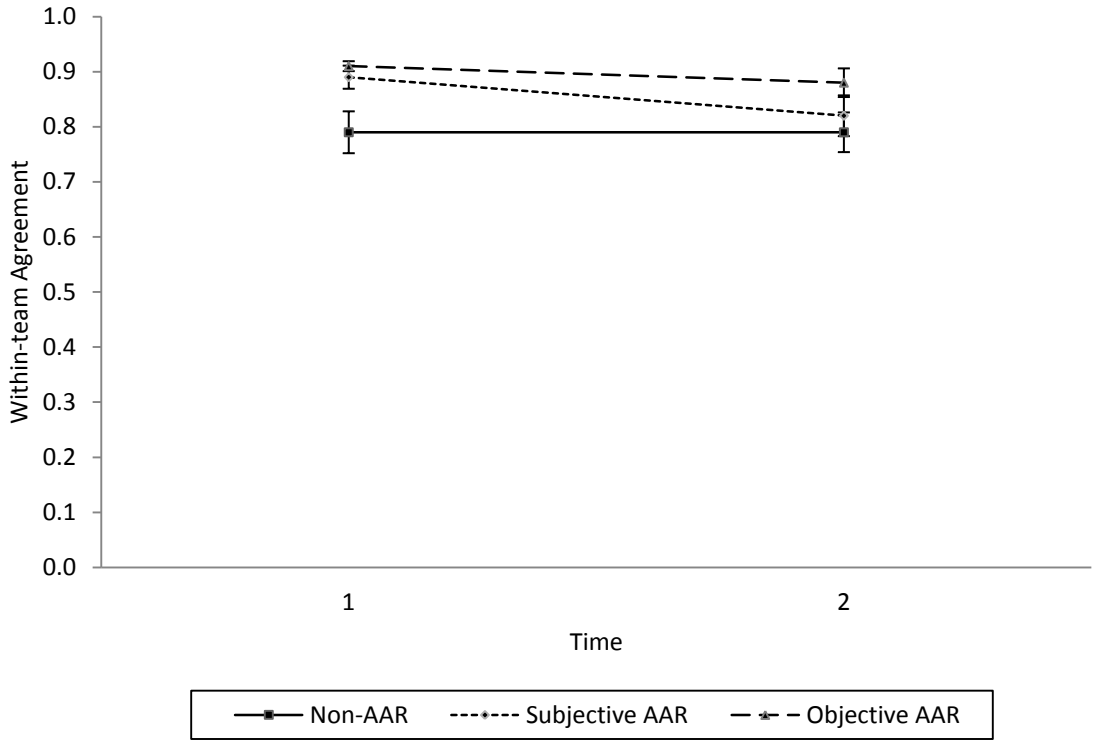


Figure 9. Within-team agreement on team-efficacy over time (by AAR condition).

DISCUSSION

The objective of the present study was to investigate the relationship between team performance and team-efficacy in the context of AARs. Specifically, the study sought to explore the role of the accuracy of performance feedback on team-efficacy—that is, on teams' ability to accurately calibrate performance and assessments of ability. A summary of the results for the research hypotheses is presented in Table 10.

Table 10

Summary of the Results of Hypothesis Tests

Hypothesis	Support
1a: Team performance will positively predict team-efficacy.	Partial
1b: The team performance and team-efficacy relationship will be moderated by the level of objectivity of an AAR. Specifically, this relationship will be stronger for teams trained using objective AARs than for teams trained using subjective AARs, and stronger for teams trained using subjective AARs than for teams trained without AARs.	Partial
2a: Team-efficacy will be better predicted by more proximal team performance than by more distal team performance.	Yes
2b: The increase in the predictive validity of team performance on team-efficacy will be greater for teams trained using objective AARs than for teams trained using subjective AARs, and greater for teams trained using subjective AARs than for teams trained without AARs.	Partial
3: Within-team agreement of team-efficacy ratings will increase over time and this increase will be stronger for teams that engage in an AAR, regardless of the level of objectivity of the AAR.	No

Note. AAR = After-action review.

Consistent with Bandura's self-regulation theory (1997) and Powers' (1973) control theory, Hypothesis 1a stated that performance would predict efficacy at the team level. In addition, consistent with views on the critical role of accurate feedback on the congruence between performance and efficacy perceptions (e.g., Bandura, 1997; Lindsley et al., 1995), Hypothesis 1b stated that the relationship between performance and efficacy at the team level would be moderated by the level of objectivity of an AAR such that this relationship would be stronger for teams trained using objective AARs than for teams trained using subjective AARs, and stronger for teams trained using subjective AARs than for teams trained without AARs. Overall, performance significantly predicted efficacy at the completion of training, but not at the beginning of training. However, the level of objectivity of an AAR did not moderate the relationship between performance and efficacy as predicted. At the completion of training, performance predicted efficacy among teams trained using objective AARs. However, performance predicted efficacy even more strongly among teams that did not participate in any AARs, and the relationship between performance and efficacy was weakest among teams trained using subjective AARs. This seemingly contradictory is discussed in latter sections of the *Discussion*.

Hypothesis 2a stated that the benefits—in terms of the degree of congruence between performance and efficacy—of AAR objectivity would increase across time. That is, efficacy would be better predicted by more proximal performance than by more distal performance at the team level. In addition, Hypothesis 2b stated that this relationship would be moderated by the level of objectivity of an AAR such that the

increase in the predictive validity of performance on efficacy would be greater for teams trained using objective AARs than for teams trained using subjective AARs, and greater for teams trained using subjective AARs than for teams trained without AARs. Overall, the predictive validity of performance on efficacy increased as performance became more proximal to efficacy. However, the level of objectivity of an AAR did not moderate the relationship between performance and efficacy across time as predicted. Specifically, the most proximal performance predicted efficacy most strongly if teams participated in either an objective AAR or no AAR. In contrast, among teams that participated in subjective AARs, the most *distal* performance predicted efficacy most strongly.

Finally, Hypothesis 3 stated that within-team agreement on efficacy ratings would increase over time and this increase would be greater for teams that engaged in AARs, regardless of the level of objectivity such AARs. Contrary to Hypothesis 3, within-team agreement on efficacy *decreased* across time. Furthermore, this decrease was greater for teams that participated in AARs than for teams that did not. In other words, AARs caused team members to become more dissimilar in their perceptions of their team's efficacy.

Theoretical Implications

Despite findings in the performance and efficacy literature (Arthur et al., 2007; Heggestad & Kanfer, 2005; Judge et al., 2007; Richard et al., 2006; Schmidt & DeShon, 2010; Shea & Howell, 2000; Vancouver et al., 2002; Vancouver et al., 2001) positing a positive relationship between performance and efficacy, the current study found an

inconsistent relationship between team performance and team-efficacy in the context of AARs and as such, may indicate an important boundary condition on the performance-efficacy relationship.

The divergence between the present findings and the extant literature on the relationship between performance and efficacy at the team level may be attributed to the inclusion of different instantiations of the AAR process in the current study. Specifically, few studies in the extant literature include AARs as part of their experimental protocols. Some researchers have manipulated task feedback (e.g., Schmidt & DeShon, 2010; Shea & Howell, 2000); however, as far as can be ascertained, even fewer researchers (e.g., Ellis et al. [2010], and Savoldelli et al. [2006], at the individual level; Villado [2008], at the team level) have provided teams the opportunity to systematically review performance in an AAR. Although AARs have received scant attention in the performance and efficacy literature, investigating the effect of AARs on the performance-efficacy relationship contributes to the understanding of the relationship between performance and efficacy over time. In particular, investigating the effect of AARs on the relationship between performance and efficacy improved our understanding of the boundary conditions under which efficacy is influenced by performance over time.

The relationship between performance and efficacy varied as a function of the type of AAR in which a team participated. For teams that trained with objective AARs, performance did not initially predict efficacy; however, at the completion of training, performance predicted efficacy much more strongly. This outcome is similar to Arthur

et al.'s (2007) finding that performance predicted efficacy as a function of time; however, Arthur et al. reported a much stronger correlation (r = .45) between performance and efficacy at an early stage of training. Moreover, participants in Arthur et al. (2007) received objective performance scores but were not afforded the opportunity to review performance in an AAR. In contrast, subjective AARs resulted in a relationship between performance and efficacy which was initially strong, but steadily declined across time. One explanation for these ambiguous findings—that efficacy became more strongly predicted by performance with objective AARs and less strongly predicted by performance with subjective AARs—may rest with the nature of subjective AAR. Participants in subjective AARs reviewed performance without the assistance of memory aids; as such, they lacked the opportunity to verify existing perceptions about performance against an objective source of information (i.e., a video replay of the performance episode) and may have succumbed to cognitive biases regarding their performance (Ellis & Davidi, 2005). These potentially inaccurate perceptions of performance may have manifested themselves in efficacy perceptions that became less congruent with actual performance. Specifically, an examination of Table 5 indicates that—when compared to objective AARs—subjective AARs had slightly (although not significantly) lower performance scores coupled with slightly (although not significantly) higher efficacy ratings.

It is also possible that without veridical data against which to verify perceptions of performance, team members also may have accepted teammates' potentially error-prone perceptions of performance. Villado (2008) alluded to this, noting anecdotally

that teams trained using subjective AARs often experienced a lack of situational awareness during training missions that would "inevitably permeate the subsequent review" (p. 96) and in some instances, teammates' incorrect shared information had a downward synergistic effect in which teammates agreed on incorrect information supported by prior incorrect information; this phenomenon was also observed during the present study. This suggests that the factual errors present during the subjective review may not have been particularly consequential and despite the subjective review being susceptible to errors, the errors present during such reviews did not hinder teams' learning of concepts and strategy development, as evidenced by their performance. It would seem that teams may have benefitted from both individual-level meta-cognitive and team-level macrocognitive activity brought about by the AAR. That is, despite making factual errors (e.g., the location of enemy positions), teams were able to discuss and develop strategies for improving their performance (e.g., how to effectively search for and destroy the enemy).

In summary, the results of the present study reveal no significant differences in the performance or efficacy levels between the subjective and objective AARs. Teams trained using subjective AARs performed as well as and had efficacy levels similar to teams trained using objective AARs, but their perceptions of efficacy were not as congruent with actual performance.

In addition to the contrasting performance-efficacy relationships seen in the subjective and objective AAR conditions, the similarity between the non- and objective AAR conditions is particularly noteworthy. Although it was hypothesized that the

efficacy ratings of teams lacking an opportunity to systematically review their performance would be least congruent with past performance, the exact opposite result was found—the strongest relationship between performance and efficacy was observed for teams that did not participate in AARs. At first glance, it appears that this finding is at odds with findings such as those reported by Schunk (1983). He found that children receiving objective feedback (i.e., a written number indicating the number of pages of math problems completed) had significantly higher self-efficacy than children in a no-feedback condition. However, teams in the non-AAR condition in the present study were not completely bereft of feedback—each team was provided an opportunity to view its performance score at the completion of a performance episode. As such, the non-AAR condition in the present study shares at least some characteristics with the objective feedback condition described by Schunk (1983). In addition, teams in the non-AAR condition (i.e., teams who did not participate in AARs) are similar to teams in both the additive and referent-shift efficacy operationalization conditions described by Arthur et al. (2007)—specifically, both lacked anything (e.g., discussion about and reflection on performance experiences) to muddle the relationship between performance and efficacy. Therefore, it is not surprising that both the current study and Arthur et al. report similar correlations between performance and subsequent efficacy (current study, r = .52; Arthur et al. (2007), r = .56 and .62 for additive and referent-shift efficacy operationalizations, respectively). Consistent with Guthrie's (1935) recency principle and Bandura's (1986) notion of progressive mastery, efficacy was predicted most strongly by the most proximal performance score and most weakly by the most distal performance score.

However, this pattern of results was observed only among teams that (a) did not participate in AARs and (b) participated in objective AARs. These findings are consistent with Bandura's (1986) views on efficacy as a cognitive self-regulatory process that is critical to progressive mastery and with Gist and Mitchell's (1992) view of efficacy as a dynamic construct that constantly changes as new information and experience are acquired. Specifically, it appears that team-efficacy is a dynamic construct and teams revised efficacy assessments in response to performance episodes.

Based on previous empirical findings (e.g., Arthur et al., 2007; Heggestad & Kanfer, 2005; Shea & Howell, 2000; Vancouver et al., 2002), it was assumed that performance would positively predict efficacy. However, in both AAR conditions, efficacy at the end of training was negatively predicted by the most distal performance and positively predicted by the most proximal performance. That is, teams that initially performed poorly finished with higher efficacy perceptions, and teams that initially performed well finished with lower efficacy perceptions. An explanation for the finding that teams with poorer initial performance ended with higher efficacy perceptions can be found in Powers (1973) control theory. Specifically, recognition of poor performance may have resulted in increased effort and application of more appropriate performance strategies in subsequent performance episodes, which resulted in increased performance, which resulted in increased efficacy.

An explanation for the finding that teams with initially higher (but not initially lower) performance ended with lower efficacy perceptions may be found in Janis' (1983) groupthink model, which Gist (1987) also invoked as a potential explanation for

unrealistically high efficacy perceptions. According to Janis, defective decision-making results from cohesive groups that collectively rationalize (i.e., group members discount or withhold derogatory information that could threaten the group's beliefs and refrain from consideration of outside information) or self-censor (i.e., group members refrain from expressing doubts from an apparent group consensus). In the present study, team members were randomly assigned to teams and there may have been little incentive to "rock the boat" and create discord during AARs by providing information regarding a team's shortcomings and opportunities for improvement. In response to this phenomenon, Cannon-Bowers, Salas, and Converse (1993) suggested that support systems that provide teams with alternative hypotheses about a situation may combat team members' desire for conformity and agreement. AARs provide such support systems, and as teams participate in successive AARs, they are provided with increased opportunities to explore alternative hypotheses, resulting in efficacy perceptions that become congruent with performance. However, in the current study, it is possible that teams did not perform as a team long enough for team members to overcome the need to rationalize or self-censor and the benefits of AARs (in the context of the congruency between performance and efficacy) were not fully realized. Similar results were observed by Shea and Howell (2000), who speculated that participants with initially high levels of performance may have become complacent, considering themselves as highly efficacious in terms of task performance and too content to motivate themselves to do better.

Although the AAR process provides a forum for information sharing (Mesmer-Magnus & DeChurch, 2009) and feedback (Villado, 2008)—both factors expected to contribute to increases in within-team agreement about team-efficacy—the opposite result was obtained. That is, teams that participated in AARs actually displayed a slight decrease in agreement. The mental model literature (cf., Cannon-Bowers et al., 1993) has generally subscribed to the notion that shared task knowledge results in team members having compatible expectations for performance. In other words, team members who share task knowledge should have more similar mental models (and hence, more accurate expectations and predictions regarding performance) than those that do not share task knowledge. It would appear, however, that this belief rests on the assumption that the information shared among team members is accurate and that team members' interpretations of this information is also error-free. In the present study, a facilitator ensured that the teams completed the steps of the AAR in the prescribed order. When a team deviated from the order, the facilitator interrupted the AAR and reminded participants to complete the AAR in the order prescribed; thus the interaction between the experimenter and participants was regimented and minimal. The facilitator did not correct participants' misinterpretations, mistakes, or errors in processing feedback about the performance episode—and such misinterpretations occurred frequently during AARs. The existence of these types of information processing errors could explain why teams that shared information (with the concomitant errors in interpreting such information) displayed decreases in within-team agreement, especially on a complex, highly interdependent task, such as the one used in the present study. This reasoning

may also explain why teams that did not participate in AARs did not show a decrease in within-team agreement. The only piece of information on which these team members could base perceptions of efficacy was the performance score given at the end of a performance episode and the lack of an opportunity to share information during an AAR may have minimized the occurrence of information processing errors and subsequent divergent perceptions of efficacy.

Practical Implications

Although the hypotheses presented were only partially supported, the findings have potentially important practical implications for team training. Specifically, designers of training programs may want to be aware of the roles that both AAR objectivity and time play in teams' perceptions of efficacy. For example, whereas subjective AARs resulted in efficacy perceptions more congruent with performance earlier in training, objective AARs resulted in efficacy perceptions more congruent with performance later in training. As such, a training program in which teams review performance without the aid of objective performance feedback at early stages of training, followed by reviews of performance aided by objective performance feedback could potentially further increase the congruence of efficacy perceptions with performance. Efficacy perceptions that are more reflective of prior performance may subsequently result in more appropriate goal-setting and increased rehearsal and effort directed toward future performance episodes. For example, consider two teams with equally poor prior performance. A team with a lower efficacy perception that is more reflective of the poor past performance may set a lower—albeit more appropriate—goal

for future performance than a team with a higher efficacy perception that is less reflective of the poor past performance. This reasoning introduces a measure of doubt to the notion that increasing efficacy is always a desirable goal (Bandura, 2012; Gist & Mitchell, 1992; Karl, O'Leary-Kelly, & Martocchio, 1993). Thus, the current findings may indicate that efficacy perceptions that are more congruent with prior performance could be more desirable (Lindsley et al., 1995) than unwarranted high efficacy perceptions that may result in overconfidence and complacency, as suggested by Gist and Mitchell (1992).

Insofar as efficacy perceptions that are more congruent with past performance are more desirable than efficacy perceptions that are less congruent with past performance, it may also be beneficial to consider the extent to which emergent or pre-existing team characteristics may influence teams' efficacy perceptions. For example, Weiner (1986) suggested that after an outcome, individuals undertake a causal search for why the outcome occurred. He proposed that individuals who perform well attribute their success to their own ability. However, individuals who do not perform well search for factors responsible for their failure. Weiner suggested that a large number of antecedents, including both one's own performance and the performance of others influence these causal explanations. In addition, Silver, Mitchell, and Gist (1995) found that self-efficacy moderated the relationship between performance and attributions—high-efficacy individuals attributed unsuccessful performance to external factors such as bad luck, but low-efficacy individuals attributed unsuccessful performance to low ability. In contrast, all individuals attributed successful performance to internal traits

such as ability. An extension of Weiner's reasoning to the team level suggests that a team's efficacy perceptions may be influenced by attributions regarding the causes of performance. For example, teams that attribute failure to external factors may be more likely to exhibit unjustifiably high efficacy perceptions. Furthermore, as teams search for causal explanations of performance, the degree to which teams tend to attribute success or failure to internal or external causes may influence the degree to which efficacy perceptions are congruent with past performance. Therefore, the degree to which an AAR provides teams the opportunity to discover the true causes of performance may be a desirable feature in the context of team training.

In addition to providing teams with a forum for review and analysis of past performance, AARs also provide a venue in which teams use their review of past performance to develop strategies for subsequent performance episodes. The cyclical nature of team functioning implies that strategies developed during an AAR affect subsequent performance and efficacy. Therefore, monitoring teams' strategy decisions may facilitate investigations into why efficacy perceptions may not be congruent with past performance. For example, consider a team that performs well, rates their efficacy as high, changes strategy for a subsequent performance episode, and then performs poorly as a result of selecting a flawed strategy. It would not be unreasonable to expect the team's efficacy perception to remain high, even in the face of poor performance. It could be argued that attributing the poor performance to a flawed strategy could be considered a specific example of attributing poor performance to an external factor. However, certain training methods such as error management training encourage

exploration, experimentation, and commission of errors (Frese, Brodbeck, Heinbockel, Mooser, Schleiffenbaum, & Thiemann, 1991). If teams were to view the selection of flawed performance strategies as an integral part of the training process and as opportunities to learn (Frese et al., 1991), and such strategy selections were monitored and tracked across time, then incongruence between performance and efficacy could be explained by changes in strategy.

Finally, subjective AAR design features may be effective in environments or for tasks where it would not be feasible to incorporate objective review systems. For example, it may not be possible to provide a veridical replay of team members' actions such as that provided during the objective AARs in the present study. However, such a lack of objective data may not hinder the effectiveness of training that uses AARs, as evidenced by higher performance scores at the end of training. Because an investigation of the effectiveness of AARs in various environments and for various tasks is critical to building a comprehensive understanding of training that uses AARs, future research should seek to identify boundary conditions that limit the effectiveness of both subjective and objective AARs. In summary, the counter-intuitive nature of these findings (i.e., that a systematic review based on subjective performance data results in teams that perform as well as—but do not realize they are capable of performing as well as—those reviewing objective performance data) warrants further investigation and serves as a cautionary note; training interventions must be submitted to empirical scrutiny—one should not rely on only their seemingly intuitive utility to justify and support their continued use.

Future Research

In the current study, both the non- and objective AAR conditions resulted in efficacy perceptions that were unrelated to performance early in training, but which were predicted well by performance at the end of training. Although future research may seek to discover the characteristics which are common to both conditions, an alternative—and perhaps more interesting—question may be, "What characteristics unique to the subjective AAR condition result in such differential results (i.e., efficacy perceptions that are predicted by performance more strongly earlier in training than at a later stage of training)?"

In addition, the conflicting results previously discussed—namely that the objectivity of an AAR results in efficacy beliefs that are differentially reflective of prior performance at different times in the training process—suggests that a combination of AAR approaches may be beneficial. That is, to increase the congruence of efficacy beliefs with prior performance both early and later in training, it could be the case that providing teams the opportunity to review performance *without* the aid of objective performance feedback early in training, followed by a transition to a situation in which teams are provided with the opportunity to review performance *with* the aid of objective performance feedback later in training could result in increasing the degree to which efficacy beliefs are reflective of prior performance to an even greater degree. Of course, when to make such a transition is a question for future empirical research.

Although teams that did not have an opportunity to review prior performance during AARs had efficacy beliefs at the end of training that were most reflective of prior

performance, it should be noted that such teams also had the lowest performance levels. Thus, another question for future research may involve the desirability of outcomes. That is, should the goal of training be to (a) maximize performance, or to (b) maximize the congruency between of teams' beliefs about their capabilities and their actual performance? The results of the present study imply that these goals may be in conflict, but it could be the case that each of these apparently conflicting goals are more desirable at different stages in a training process. For example, it may be that maximizing accurate assessment of abilities is desirable earlier in training, while maximizing performance attainment is desirable at the completion of training.

In addition, although the current study did not find that objective AARs resulted in a stronger relationship between performance and efficacy, future research should explore situations in which objective and subjective AARs are more or less valuable. For example, Gist and Mitchell (1992) theorized that when teams are experienced with a task, it is less likely that a detailed task and resource analysis will form the basis for efficacy beliefs and more likely that efficacy beliefs will be based on past performance. In addition, Gist and Mitchell also suggest that—in general—teams with higher levels of task experience should have efficacy beliefs that are more reflective of past performance. Thus, a team's level of task experience could be explored as a potentially moderating factor.

Schneider (1985) proposed that the cause-and-effect relationships between cues, acts, and work products are generally more difficult to discern in complex tasks—such as the one used in the present study—because complex tasks typically have a large

number of relationships between the cues and acts that result in work products. In addition, the relationships themselves are often dynamic, making both learning from and self-assessment of performance even more difficult during the performance of complex tasks. In contrast to complex tasks, simple tasks generally display relatively fewer cause-and-effect relationships between cues, acts, and work products. Because AARs aid teams in specifically identifying critical behavior-outcome relationships, they should have a greater effect on complex than on simple tasks. As such, task complexity may also serve as a potential boundary condition in the context of the relationship between performance and efficacy.

Finally, it is unclear whether the differences observed in the performance-efficacy relationships (between AAR conditions) are due to the objectivity of feedback and observational learning present in objective AARs or to experience with the task itself. Specifically, teams in the objective AAR condition—by virtue of interaction with the video replay of performance—received a greater amount of task exposure than those in the subjective AAR condition. Although it is possible that this increased task exposure could be responsible for the stronger relationship between performance and efficacy observed in the objective than the subjective AAR condition, it is unclear why teams who received no feedback—and were therefore exposed to the task for the same period of time as those in the subjective AAR condition—exhibited a performance-efficacy relationship of the same magnitude as that observed in the objective AAR condition. It is possible that the increased task exposure results in increased levels of performance (as observed in the objective AAR condition), but the rationalization and

self-censorship present in AARs result in efficacy perceptions that are less congruent with performance (as observed in the subjective AAR condition).

Limitations

There are some limitations with the present study that are worth noting but these potential limitations may also yield fruitful lines of future research. First, a major limitation concerns the criteria used to define the end of training. Participants trained for a specific amount of time (i.e., 45 min of individual training and 90 min of team training) rather than to a specified level of performance (e.g., three errorless trials, or asymptote). This is noteworthy in that different criteria represent different dimensions of skill acquisition, retention, and transfer (Arthur, Bennett, Stanush, & McNelly, 1998; Schmidt & Björk, 1992). In their quantitative review of the skill decay literature, Arthur et al. noted that several definitions of performance have been used throughout the training literature to identify the end of skill acquisition. In particular, they noted that two methods of measuring skill acquisition have generally been used in the extant literature—namely (a) how much is trained in a given amount of time and (b) how long it takes to train a given amount of material. In the current study, the end of acquisition was defined as the former. How skill acquisition is defined is not trivial because the degree to which an AAR's level of objectivity influences the congruence between efficacy and prior performance when acquisition is defined by a fixed amount of time may differ from its degree of influence when acquisition is defined by a specified level of performance. The present study demonstrated that subjective and objective AARs influence the relationship between performance and efficacy differentially at the

beginning and end of training. Because the effectiveness of training interventions varies as a function of the phase of skill acquisition (Arthur et al., 1998), it may be worth exploring the influence of an AAR's level of objectivity in later stages of skill acquisition, retention, or reacquisition to obtain a more complete understanding of the influence of AARs.

A second limitation concerns the effect of familiarity on the relationship between efficacy and performance. Teams in the non-AAR condition completed filler tasks instead of participating in AARs. Although this ensured equal spacing between performance sessions, it did not ensure that team members interacted with each other for equal periods of time. Therefore, the higher performance levels exhibited by teams in the AAR conditions—as compared to teams in the non-AAR condition—could be explained by cohesion related to familiarity (Gully, Devine, & Whitney, 1995). However, such difference in familiarity does not explain the difference observed in the performance-efficacy relationships between the subjective and objective AAR conditions. In both conditions, team members spent equal amounts of time interacting with each other, and as such, likely developed similar levels of familiarity. However, the objective AAR condition resulted in stronger performance-efficacy relationships than the subjective AAR condition. Although the nature of the filler task used in the current study was such that it did not ensure equal levels of familiarity across AAR conditions, nothing precludes the use of such a task (e.g., a team discussion of an topic unrelated to the performance task) in future research. Although differing levels of familiarity are inherently a characteristic of AAR research involving the comparison of treatment (i.e.

AAR), and control (i.e., non-AAR) conditions, future AAR research could benefit from development and implementation of appropriate filler tasks that control for familiarity.

A third limitation concerns the degree to which teams accurately processed performance feedback during the AARs. In running the study protocols, two experimenters served as facilitators for the AAR sessions. As facilitators, the experimenters ensured that participants completed the steps of the AAR in the prescribed order. When a team deviated from the order, the experimenter interrupted the AAR and reminded participants to complete the AAR in the order prescribed. Thus, facilitators only corrected teams' errors in adhering to the AAR process and did not correct teams' errors in interpreting information during the AARs. Consequently, AARs were often rife with misinterpretations of the actions taken during prior performance episodes. For example, in both subjective *and* objective AARs, teams displayed errors of fact (e.g.., the location of friendly and enemy tanks during critical situations, the identity of a tank engaging or being engaged by an enemy, the identity of a target, and the results of shots fired). Given these errors in interpreting even objective performance feedback—in the form of a video replay of performance—it would be worthwhile to examine the effectiveness of AARs with regard to how the review is led or facilitated (e.g., expert-led versus leaderless).

Finally, the degree of information available to teams participating in objective AARs—coupled with the time limit within which AARs were conducted—may have impacted the degree to which definitive conclusions can be reached regarding the influence of feedback accuracy on the relationship between performance and efficacy.

Teams participating in objective AARs were exposed to a greater amount of information during the performance review. This may have resulted in teams that spent—compared to those in the subjective AAR condition—a greater percentage of AAR time reviewing specific performance incidents and a smaller percentage of AAR time spent devoted to higher-level strategic decisions for subsequent performance trials. Thus, any systematic differences in the degree of information availability and time constraints experienced by teams may have resulted in qualitatively different psychological processes influencing teams' performance-efficacy relationships; for example, the relationship between performance and efficacy may have been influenced more by task experience in the objective AAR condition and more by better strategic decision-making in the subjective AAR. Given the presence of these potential alternative explanations for the present results, it may be worthwhile to investigate the extent to which AAR design characteristics are conducive to different psychological processes.

Conclusion

The use of AARs in non-military organizations has increased dramatically (Zakay, Ellis, & Shevaliski, 2004). Yet, despite the prevalence of AARs in both military and non-military settings, researchers have only recently begun to empirically investigate the AAR. This study sought to explore the relationship between performance and efficacy at the team level in the context of AARs. As such, the current study advances the training literature by empirically investigating how the accuracy of feedback available to trainees during AARs influences perceptions of team-efficacy. In addition, it provides insight into how the predictive validity of team performance on

team-efficacy changes as a function of both time and feedback accuracy. Finally, this research provides preliminary guidance to researchers and practitioners in more effectively integrating AAR design characteristics into team training environments.

The results of this study suggest that the accuracy of performance feedback provided to teams during AARs is differentially effective in increasing the congruence between performance and efficacy across time. Whereas the mechanisms influencing this differential effectiveness across time remain unclear, it does appear that providing accurate and objective performance feedback to teams facilitates accomplishing the simultaneous goals of increasing both (a) a team's level of performance and (b) the degree to which efficacy judgments are reflective of past performance.

REFERENCES

Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied, 1,* 270-303.

Alexander, L. T., Kepner, C. H., & Tregoe, B. B. (1962). The effectiveness of knowledge of results in a system-training program. *Journal of Applied Psychology, 46,* 202-211.

Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research, 61,* 315-343.

Ancona, D. G., Goodman, P. S., Lawrence, B. S., & Tushman, M. L. (2001). Time: A new research lens. *Academy of Management Review, 26,* 645-663.

Ancona, D. G., Okhuysen, G. A., & Perlow, L. A. (2001). Taking time to integrate temporal research. *Academy of Management Review, 26,* 512-529.

Arthur, W. Jr., Bell, S. T., & Edwards, B. D. (2007). A longitudinal examination of the comparative criterion-related validity of additive and referent-shift consensus operationalizations of team-efficacy. *Organizational Research Methods, 10,* 35-58.

Arthur, W. Jr., Bennett, W. Jr., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative analysis. *Human Performance, 11,* 57-101.

Arthur, W. Jr., Edwards, B. D., Bell, S. T., Villado, A. J., & Bennett, W. Jr. (2005).

Team task analysis: Identifying tasks and jobs that are team based. *Human Factors, 47,* 654-659.

Arthur, W. Jr., Glaze, R. M., Bhupatkar, A., Villado, A. J., Bennett, W. Jr., & Rowe, L. (in press). Team task analysis: Differentiating between tasks using team-relatedness and team workflow as metrics of team task interdependence. *Human Factors.*

Arthur, W. Jr., Jarrett, S. M., Glaze, R. M., Schurig, I., Munoz, G. J., McDonald, J. N., Naber, A. M., & Villado, A. J. (2011). *An investigation of factors that moderate the effectiveness of after-action reviews: Objective and subjective after-action reviews in co-located and distributed training environments.* Contract No. FA8650-05-D-6502, L-3 Communications Corporation, Link Training and Simulations Systems, Mesa, AZ; and Department of Defense, USAF Air Force Research Laboratory (AFRL), Warfighter Training Research Division, Mesa, AZ.

Ashford, S. J. (1989). Self-assessments in organizations: A literature review and integration. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (pp. 133-174). Greenwich, CT: JAI Press.

Baker, D. F. (2001). The development of collective efficacy in small task groups. *Small Group Research, 32,* 451-474.

Balzer, W. K., Doherty, M. E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin, 106,* 410-433.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change.

Psychological Review, 84, 191-215.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37,* 122-147.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science, 9,* 75-78.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management, 38,* 9-44.

Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88,* 87-99.

Barclay, D., Higgins, C., & Thompson, R. (1995). The partial least squares (PLS) approach to causal modeling: Personal computer adoption and use as an illustration. *Technology Studies, 2,* 285-324.

Bartko, J. J. (1976). On various intraclass reliability coefficients. *Psychological Bulletin, 83,* 762-765.

Borkovec, T. D. (1978). Self-efficacy: Cause or reflection of behavioral change. *Advances in Behavioral Research and Therapy, 1,* 163-170.

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica, 45,* 223-241.

Cannon-Bowers, J. A, Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision making. In N. J. Castellan, Jr. (Ed.), *Individual and group*

decision making: Current issues (pp. 221-246). Hillsdale, NJ: Erlbaum.

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83,* 234-246.

Chen, G., & Kanfer, R. (2006). Toward a systems theory of motivated behavior in work teams. *Research in Organizational Behavior, 27,* 223-267.

Chen, G., Kanfer, R., DeShon, R. P., Mathieu, J. E., & Kozlowski, S. W. J. (2009). The motivating potential of teams: Test and extension of Chen and Kanfer's (2006) cross-level model of motivation in teams. *Organizational Behavior and Human Decision Processes, 110,* 45-55.

Chen, G., Thomas, B., & Wallace, J. C. (2005). A multilevel examination of the relationships among training outcomes, mediating regulatory processes, and adaptive performance. *Journal of Applied Psychology, 90,* 827-841.

Chi, M. T. H., de Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439-477.

Chin, W. W. (2000). Frequently asked questions: Partial least squares and PLS-Graph. Retrieved November 15, 2011 from http://disc-nt.cba.uh.edu/chin/plsfaq/plsfaq.htm

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Corrigan, P. W. (1990). Theoretical weakness in behavior theory is no more than statistical variance: Response to Christina Lee. *Journal of Behavior Therapy and Experimental Psychiatry, 21,* 141-142.

DeNisi, A. S., Robbins, T., & Cafferty, T. P. (1989). Organization of information used for performance appraisals: Role of diary-keeping. *Journal of Applied Psychology, 74,* 124-129.

DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Weichmann, D. (2004). A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. *Journal of Applied Psychology, 89,* 1035-1056.

Devine, D. J., Clayton, L. D., Philips, J. L., Dunford, B. B., & Melner, S. B. (1999). Teams in organizations prevalence, characteristics, and effectiveness. *Small Groups Research, 30,* 678-711.

Earley, P. C., & Lituchy, T. R. (1991). Delineating goal and efficacy effects: A test of three models. *Journal of Applied Psychology, 76,* 81-98.

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

Ellis, S., Ganzach, Y., Castle, E., & Sekely, G. (2010). The effect of filmed versus personal after-event reviews on task performance: The mediating and moderating role of self-efficacy. *Journal of Applied Psychology, 95,* 122-131.

Ellis, S., Mendel, R, & Nir, M. (2006). Learning from successful and failed experiences: The moderating role of kind of after-event review. *Journal of Applied Psychology, 91,* 669-680.

eSim Games. (2007). *Steel Beasts Pro PE ver. 2.370* [Computer software]. Mountain View, CA: Author.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

Faulk, F. R., & Miller, N. B. (1992). *A primer for soft modeling.* Akron, OH: Akron Press.

Feltz, D. L. (1988). Self-confidence and sports performance. *Exercise and Sport Sciences Reviews, 16*, 423-458.

Fischhoff, B. (1982). For those condemned to study the past: Heuristics and biases in hindsight. In D. Kahanman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 335-352). Cambridge, England: Cambridge University Press.

Fornell, C. (Ed.). (1982). *A second generation of multivariate analysis, Vol. I: Methods.* New York: Praeger.

Fornell, C., & Bookstein, F. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research, 19*, 440-452.

Frese, M., Brodbeck, F. C., Heinbokel, T., Mooser, C., Schleiffenbaum, E., & Thiemann, P. (1991). Errors in training computer skills: On the positive function of errors. *Human–Computer Interaction, 6*, 77-93.

George, J. M., & Jones, G. R. (2000). The role of time in theory and theory building. *Journal of Management, 26*, 657-684.

Gibson, C. B. (2001). Me and us: Differential relationships among goal-setting training, afficacy and effectiveness at the individual and team level. *Journal of Organizational Behavior, 22,* 789-808.

Gibson, C. B., Randel, A., & Earley, P. C. (2000). Understanding group efficacy: An empirical test of multiple assessment methods. *Group & Organization Management, 25,* 67-97.

Gist, M. E. (1987). Self-efficacy: Implications for organizational behavior and human resource management. *Academy of Management Review, 12,* 472-485.

Gist, M.E., & Mitchell, T.R. 1992. Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management Review, 17,* 183-211.

Gully, S. M., Devine, D. J., & Whitney, D. J. (1995). A meta-analysis of cohesion and performance: Effects of level of analysis and task interdependence. *Small Group Research, 26,* 497-520.

Gully, S. M., Incalcaterra, K. A., Joshi, A., & Beaubien, J. M. (2002). A meta-analysis of team-efficacy, potency, and performance: Interdependence and level of analysis as moderators of observed relationships. *Journal of Applied Psychology, 87,* 819-832.

Guthrie, E. R. (1935). The psychology of learning. New York: Harper & Row.

Hawkins, R. M. (1992). Self-efficacy: A predictor but not a cause of behavior. *Journal of Behavior Therapy and Experimental Psychiatry, 23,* 251-256.

Heggestad, E. D., & Kanfer, R. (2005). The predictive validity of self-efficacy in training performance: Little more than past performance. *Journal of Experimental Psychology: Applied, 11,* 84-97.

Henshel, R. L. (1976). *On the future of social prediction.* Indianapolis, IN: Bobbs-Merrill.

Hodges, L., & Carron, A. V. (1992). Collective efficacy and group performance. *International Journal of Sports Psychology, 23,* 48-59.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64,* 349-371.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group reliability with and without response bias. *Journal of Applied Psychology, 69,* 85-98.

Janis, I. L. (1983). *Groupthink: Psychological studies of policy decisions and fiascoes.* Boston, MA: Houghton-Mifflin.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Users' reference guide.* Chicago, IL: Scientific Software International, Inc.

Judge, T. A., Jackson, C., Shaw, J. C., Scott, B. A., & Rich, B. L. (2007). Self-efficacy and work-related performance: The integral role of individual differences. *Journal of Applied Psychology, 92,* 107-127.

Jung, D. I., & Sosik, J. J. (1999). Effects of group characteristics on work group performance: A longitudinal investigation. *Group Dynamics: Theory, Research, and Practice, 3,* 279-290.

Jung, D. I., & Sosik, J. J. (2003). Group potency and collective efficacy: Examining their predictive validity, level of analysis, and effects of performance feedback on future group performance. *Group and Organization Management, 28,* 366-391.

Karl, K. A., O'Leary-Kelly, Martocchio, J. J. (1993). The impact of feedback and self-efficacy on performance in training. *Journal of Organizational Behavior, 14,* 379-394.

Katz-Navon, T. Y., & Erez, M. (2005). When collective- and self-efficacy affect team performance: The role of task interdependence. *Small Group Research, 36,* 437-465.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119,* 254-284.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 349-370.

Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest, 7,* 77-124.

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3-90). San Francisco, CA: Jossey-Bass.

Lee, C. (1989). Theoretical weakness lead to practical problems: The example of self-efficacy theory. *Journal of Behavior Therapy and Experimental Psychiatry, 20,* 115-123.

Lindsley, D. H., Brass, D. J., & Thomas, J. B. (1995). Efficacy-performance spirals: A multilevel perspective. *Academy of Management Review, 20,* 645-678.

Little, B. L., & Madigan, R. M. (1997). The relationship between collective efficacy and performance in manufacturing work teams. *Small Group Research, 28,* 517-534.

Locke, E.A., Frederick, E., Lee, C., & Bobko, P. (1984). Effect of self-efficacy, goals and task strategies on task performance. *Journal of Applied Psychology, 69,* 241-251.

March, J. G. (1976). The technology of foolishness. In J. G. March & J. P. Olsen (Eds.), *Ambiguity and choice in organizations* (pp. 69-81). Bergen, Norway: Universitestsforlaget.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A conceptual framework and taxonomy of team processes. *Academy of Management Review, 26,* 356-376.

Masuch, M. (1985). Vicious circles in organizations. *Administrative Science Quarterly, 30,* 14-23.

Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology, 94,* 535-546.

Middendorf, C. H., & Macan, T. H. (2002). Note-taking in the employment interview: Effects on recall and judgments. *Journal of Applied Psychology, 87,* 293-303.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance.

Journal of Abnormal Psychology, 110, 40-48.

Mischel, L. J., & Northcraft, G. B. (1997). “I think we can, I think we can . . .”: The role of efficacy beliefs in group and team effectiveness. In B. Markovsky & M. J. Lovaglia (Eds.), *Advances in Group Processes* (Vol. 14, pp. 177-197). Greenwich, CT: JAI Press.

Mitchell, T. R., & James, L. R. (2001). Building better theory: Time and the specification of when things happen. *Academy of Management Review, 26,* 530-547.

Murphy, K. R. (1991). Criterion issues in performance appraisal research: Behavioral accuracy versus classification accuracy. *Organizational Behavior and Human Decision Processes, 50,* 45-50.

Murphy, K. R., & Balzer, W. K. (1986). Systematic distortions in memory-based ratings and performance evaluations: Consequences for rating accuracy. *Journal of Applied Psychology, 71,* 39-44.

Muthen, L. K. & Muthen, B. O. (1998). Mplus User's Guide. Los Angeles: Muthen & Muthen.

Powers, W. T. (1973). *Behavior: The control of perception.* Chicago: Aldine.

Powers, W. T. (1991). Commentary on Bandura’s “human agency.” *American Psychologist, 46,* 151-153.

Prussia, G. E., & Kinicki, A. J. (1996). A motivational investigation of group effectiveness using social-cognitive theory. *Journal of Applied Psychology, 81,* 187-198.

Richard, E. M., Diefendorff, J. M., & Martin, J. H. (2006). Revisiting the within-person self-efficacy and performance relation. *Human Performance, 19,* 67-87.

Rigdon, E. E., Ringle, C. M., & Sarstedt, M. (2010). Structural equation modeling of heterogeneous data with partial least squares. *Review of Marketing Research, 7,* 255-296.

Ringle, C. M., Wende, S., & Will, A. (2005). *SmartPLS 2.0 (M3) beta* [Computer software]. Accessed on November 1, 2011. Hamburg, Germany.

Ryan, A. M., Daum, D., Bauman, T., Grisez, M., Mattimore, K., Nalodka, & McCormick, S. (1995). Direct, indirect, and controlled observation and rating accuracy. *Journal of Applied Psychology, 80,* 664-670.

Sanchez, J. I., & De La Torre, P. (1996). A second look at the relationship between rating and behavioral accuracy in performance appraisal. *Journal of Applied Psychology, 81,* 3-10.

Savoldelli, G. L., Naik, V. N., Park, J., Joo, H. S., Chow, R., & Hamstra, S. J. (2006). Value of debriefing during simulated crisis management: Oral versus video-assisted oral feedback. *Anesthesiology, 105,* 279-285.

Schmidt, A. M., & DeShon, R. P. (2010). The moderating effects of performance ambiguity on the relationship between self-efficacy and performance. *Journal of Applied Psychology, 95,* 572-581.

Schmidt, R. A., & Björk, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207-217.

Schneider, W. (1985). Training high–performance skills: Fallacies and guidelines. *Human Factors, 27,* 285-300.

Schunk, D. H. (1983). Progress self-monitoring: Effects of children's self-efficacy and achievement. *Journal of Experimental Education, 51,* 89-93.

Schurig, I., Jarrett, S., Arthur, W. Jr., Glaze, R. M., & Schurig, M. (2011). *The effectiveness of after-action reviews as a training method: A meta-analysis.* Unpublished manuscript, College Station, TX: Texas A&M University.

Shea, C. M., & Howell, J. M. (2000). Efficacy-performance spirals: An empirical test. *Journal of Management, 26,* 791-812.

Silver, W. S., Mitchell, T. R., & Gist, M. E. (1995). Responses to successful and unsuccessful performance: The moderating effect of self–efficacy on the relationship between performance and attributions. *Organizational Behavior and Human Decision Processes, 62,* 286-299.

Sitkin, S. B. (1992). Learning through failure: The study of small losses. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (pp. 231-266). Greenwich, CT: JAI Press.

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.

Smith-Jentsch, K. A., Zeisig, R. L., McPherson, J., & Acton, B. (1998). Team dimensional training: A strategy for guided team self-correction. In J. A. Cannon-Bowers & E. Salas (Eds.), *Decision making under stress: Implications*

for individual and team training (pp. 271-297). Washington, DC: American Psychological Association.

Sosik, J. J., Kahai, S. S., & Piovoso, M. J. (2009). Silver bullet or voodoo statistics?: A primer for using the partial least squares data analytic technique in group and organization research. *Group & Organization Management, 34,* 5-36.

Stajkovic, A., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124,* 240-261.

Sturman, M. C., Cheramie, R. A., & Cashen, L. H. (2005). The impact of job complexity and performance measurement on the temporal consistency, stability, and test-retest reliability of employee job performance ratings. *Journal of Applied Psychology, 90,* 269-283.

United States Army (1993). *A leader's guide to after-action reviews* (Training Circular 25-20), Washington, DC.

Vancouver, J. B., & Kendall, L. N. (2006). When self-efficacy negatively relates to motivation and performance in a learning context. *Journal of Applied Psychology, 91,* 1146-1153.

Vancouver, J. B., More, K. M., & Yoder, R. J. (2008). Self-efficacy and resource allocation: Support for a discontinuous model. *Journal of Applied Psychology, 93,* 35-47.

Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology, 87,* 506-516.

Vancouver, J. B., Thompson, C. M., & Williams, A. A. (2001). The changing signs in the relationships among self-efficacy, personal goals, and performance. *Journal of Applied Psychology, 86*, 605-620.

Villado, A. J. (2008). *The after-action-review training approach: An integrative framework and empirical investigation. (*Doctoral dissertation). Retrieved from Proquest Dissertations & Theses. (Order Number 3333780)

Villado, A. J., & Arthur, W. Jr. (2011). *An empirical investigation of the effectiveness of subjective and objective after-action reviews in a complex task training environment.* Manuscript submitted for publication.

Weick, K. E. (1979). *The social psychology of organizing.* New York: Addison-Wesley.

Weiner, B. (1986). *An attributional theory of motivation and emotion.* New York: Springer-Verlag.

Williams, K. J., DeNisi, A. S., Meglino, B. M., & Cafferty, T. P. (1986). Initial decisions and subsequent performance ratings. *Journal of Applied Psychology, 71,* 189-195.

Woehr, D. J., & Feldman, J. (1993). Processing objective and question order effects on the causal relation between memory and judgment in performance appraisal: The tip of the iceberg. *Journal of Applied Psychology, 78,* 232-241.

Wold, H. 1985. Systems analysis by partial least squares. In Nijkamp, P., Leitner, H. & Wrigley, N. (Eds.), *Measuring the unmeasurable* (pp. 221-251). Boston, MA: Martinus Nijhoff.

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

Wood, R., & Bandura, A. (1989). Social cognitive theory of organizational management. *Academy of Management Review, 14*, 361-384.

Zaccaro, S. J., Blair, V., Peterson, C., & Zazanis, M. (1995). Collective efficacy. In J. E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment: Theory research, and application* (pp. 305-328). New York: Plenum.

Zakay, D., Ellis, S., & Shevaliski, M. (2004). Outcomes' values and early warning indications as determinates of willingness to learn from experience. *Experimental Psychology, 51*, 150-157.

APPENDIX A

TEAM-EFFICACY MEASURE

Team-Efficacy

Research ID: _____

Please read each of the statements listed below and mark the response that best indicates the extent to which you agree with each statement

①	②	③	④	⑤
Strongly Disagree	**Disagree**	**Neither Agree nor Disagree**	**Agree**	**Strongly Agree**

Rate each of the following statements to indicate the extent to which they are descriptive of **YOUR** opinions of your **PLATOON**.

1.	I feel confident in **my platoon's** ability to perform well on Steel Beasts.	① ② ③ ④ ⑤
2.	I think **my platoon** can meet the challenges of Steel Beasts.	① ② ③ ④ ⑤
3.	I know **my platoon** can achieve good scores on Steel Beasts.	① ② ③ ④ ⑤
4.	I know **my platoon** can master Steel Beasts.	① ② ③ ④ ⑤
5.	I do **NOT** think Steel Beasts is something that **my platoon** will become good at.	① ② ③ ④ ⑤
6.	I am confident that **my platoon** has what it takes to perform well on Steel Beasts.	① ② ③ ④ ⑤

APPENDIX B

DEMOGRAPHICS MEASURE

Demographics

Please read each of the statements listed below and respond in a manner that best describes you.

Research ID: _____

Age: _____ Race: _____

Sex: ☐ Male ☐ Female Major: _____

Occupation: _____

Highest education earned (check one):

Doctoral Degree	☐ Completed	☐ In-Progress			
Master's Degree	☐ Completed	☐ In-Progress			
Bachelor's Degree	☐ Completed	☐ Senior	☐ Junior	☐ Sophomore	☐ Freshman
Associates Degree	☐ Completed	☐ In-Progress			
Technical/Vocational	☐ Completed	☐ In-Progress			
High School	☐ Completed	☐ 12th grade	☐ 11th grade	☐ 10th grade	☐ 9th grade

APPENDIX C

VIDEO GAME EXPERIENCE MEASURE

Video Game Experience

Research ID: _____

Please answer each of the following questions as honestly and truthfully as possible.

1. How much time per week do you currently spend playing video/computer games?

_____ hours per week

2. During periods in your life when you consistently played video/computer games, approximately how many hours per week did you spend playing video games?

_____ hours per week

3. How long has it been since you played video/computer games with the consistency reported in Question 2?
Please use numbers, not check marks, to answer this item.

_____ years _____ months _____ days _____ hours

4. When you were in high school (age 14-17 years), approximately how many hours per week did you play video/computer games?

_____ hours per week

5. When you were in middle school (age 11-13 years), approximately how many hours per week did you play video/computer games?

_____ hours per week

6. When you were in elementary school (age 6-10 years), approximately how many hours per week did you play video/computer games?

_____ hours per week

7. Generally, what is your playing ability regarding video/computer games? (check one)
☐ Novice ☐ Average ☐ Expert

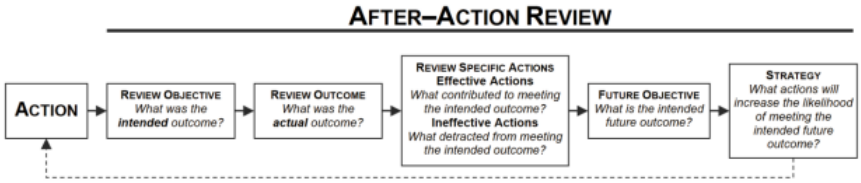
8. What is your playing ability on *Steel Beasts*™? (check one)
☐ Never Played ☐ Average ☐ Expert

APPENDIX D

AFTER-ACTION REVIEW FORM

AFTER–ACTION REVIEW FORM

An after–action review (AAR) is a systematic review of trainees' performance after a recently completed task or event. An AAR allows trainees to discover for themselves what happened, why it happened, and how to sustain strengths and improve on weaknesses. During the AAR, you will be asked to answer specific questions presented in the figure below.



The AAR is a professional discussion. Therefore, it important that you follow the guidelines listed below while conducting the AAR so that each AAR remains professional and productive. Specifically, you and you platoon should strive to:

1. **Avoid assigning blame**
2. **Focus on actions or behaviors, NOT the person**
3. **Avoid generalizations**
4. **Avoid dwelling on issues unrelated to the discussion or the mission**
5. **Participate; everyone should participate when able**

You have 10 minutes to complete each AAR, so you must work quickly, but the importance should be thoroughness and not speed. Your proctor will guide you through the AAR, however, it is the job of you and your platoon members to provide answers to the questions posed.

AFTER–ACTION REVIEW FORM

REVIEW OBJECTIVE
What was the intended outcome?

REVIEW OUTCOME
What were the exact steps you took in performing the previous missions? Did they help you to reach the mission goal? Why or Why not?

FUTURE OBJECTIVE
What is your intended future outcome for the next mission?

STRATEGY
What actions would you like use in the future missions that will help you reach the intended outcome?

APPENDIX E

EVIDENCE FOR COLLAPSING TEAMS ACROSS GEOGRAPHIC TRAINING LOCATION

The data analyzed in the current study were collected as part of a larger data collection effort in which the comparative effectiveness of AARs in co-located and distributed training environments was investigated. As such, teams were randomly assigned to AAR conditions (i.e., non-, subjective, and objective AAR) and were also randomly assigned to either a geographically co-located or distributed training environment. Because the geographic distribution of team members was not of substantive interest in this study, teams were not differentiated on the basis the geographic distribution of team members. However, before collapsing geographically co-located and distributed teams into a single group, a series of analyses of covariance (ANCOVAs) were conducted to determine whether the interaction between AAR condition and geographic training location significantly predicted any variance in the variables of interest in this study (i.e., team performance, team-efficacy, or within-team agreement about team-efficacy). The results of this series of analyses are presented in Table E1 and indicate that none of the variables of interest in the current study (i.e., team performance, team-efficacy, or within-team agreement about team-efficacy) differed on the basis of the interaction between AAR condition and geographic training location. In addition, individuals were randomly assigned to teams, and teams were randomly assigned to both AAR condition and geographic training location—all factors serving to mitigate the presence of preexisting group differences (Miller & Chapman, 2001). On

the basis of the aforementioned evidence and arguments, teams in co-located and distributed training environments were collapsed within each AAR condition.

Table E1

Means, Standard Deviations, and F Values by AAR Condition and Geographic Training Location

	Non-AAR				Subjective AAR				Objective AAR				
	Colocated (N = 23)		Distributed (N = 20)		Colocated (N = 20)		Distributed (N = 20)		Colocated (N = 20)		Distributed (N = 20)		AAR condition × geographic training location interaction F value
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
Perf 1	28.32	5.09	29.19	6.57	28.19	6.92	25.44	6.79	28.38	4.37	28.00	7.91	0.86
Eff 1	3.06	0.45	3.25	0.27	3.44	0.33	3.49	0.58	3.40	0.47	3.29	0.51	1.22
$r_{wg(j)}$ 1	.84	0.13	.73	0.34	.90	0.13	.88	0.14	.92	0.06	.91	0.05	1.02
Perf 2	31.74	9.96	36.25	8.23	35.81	9.05	35.88	9.06	36.56	9.61	38.06	8.27	0.65
Perf 3	33.53	9.93	39.06	9.34	41.38	9.36	40.56	10.47	41.25	13.78	41.06	8.65	1.19
Eff 3	3.10	0.53	3.52	0.46	3.65	0.34	3.79	0.44	3.61	0.61	3.67	0.41	1.68
$r_{wg(j)}$ 3	.79	0.26	.78	0.21	.87	0.23	.78	0.24	.87	0.21	.89	0.10	0.56

Note. AAR = After-action review; Perf 1 = performance at Time 1 (i.e., the mean of team performance scores on Test Missions 1 and 2); Eff 1 = team-efficacy at Time 1 (i.e., after Test Mission 2); $r_{wg(j)}$ 1 = within-team agreement on team-efficacy at Time 1 (i.e., after Test Mission 2); Perf 2 = performance at Time 2 (i.e., the mean of team performance scores on Test Missions 3 and 4; Perf 3 = performance at Time 3 (i.e., the mean of team performance scores on Test Missions 5 and 6; Eff 3 = team-efficacy at Time 3 (i.e., after Test Mission 6). The range of potential performance scores was from -50 to 100. For all F values in the rightmost column, df = 2,117; no F values were statistically significant at an alpha level of .05

APPENDIX F

COMPARISON OF ANALYSES CONDUCTED USING SMARTPLS AND MPLUS

In order to compare the results of the variance-based PLS approach for testing Hypotheses 1a, 1b, 2a, and 2b—conducted using SmartPLS (Ringle et al., 2005)—with a covariance-based SEM approach, the analyses conducted to test Hypotheses 1a, 1b, 2a, and 2b were replicated using Mplus (Muthen & Muthen, 1998). When teams were collapsed across AAR conditions, Mplus yielded path coefficients nearly identical to those found using SmartPLS, as seen in Table F1. Specifically, Table F1 shows that of the 56 path weights estimated, all but 9 were identical in SmartPLS and Mplus, and these 9 differed by no more than .01). However, Table F1 reveals differences in the standard errors of the estimates of path coefficients between the variance-based (SmartPLS) and covariance-based (Mplus) SEM approaches. When all AAR conditions were collapsed into a single condition (with a sample size of $N = 123$), the standard errors generated without bootstrapping (i.e., in Mplus) were consistently lower than those generated via bootstrapping in SmartPLS. In contrast, when each AAR condition was modeled separately (with smaller sample sizes of *N*s of 43, 40, and 40 for the non-, subjective, and objective AAR conditions, respectively), the standard errors generated without bootstrapping (i.e., in Mplus) were consistently higher than those generated via bootstrapping in SmartPLS.

Table F1

Comparison of Path Coefficients and Standard Errors

	Non-AAR Condition				Subjective AAR Condition				Objective AAR Condition				Overall			
	SmartPLS		Mplus		SmartPLS		Mplus		SmartPLS		Mplus		SmartPLS		Mplus	
Path	β	*SE*	β	*SE*	β	*SE*	β	*SE*	β	*SE*	β	*SE*	β	*SE*	β	*SE*
Perf1 → Perf2	.26	0.10	.26	0.14	.30	0.08	.30	0.14	.07	0.08	.07	0.16	.21	0.08	.21	0.07
Perf1 → TE1	.05	0.09	.05	0.15	.34	0.07	.34	0.13	.00	0.11	.00	0.17	.13	0.10	.13	0.09
Perf1 → TE3	.06	0.06	.06	0.11	-.29	0.11	-.29	0.12	-.15	0.08	-.15	0.13	-.11	0.09	-.11	0.08
Perf2 → Perf3	.45	0.09	.45	0.12	.45	0.08	.45	0.15	.16	0.10	.16	0.15	.32	0.09	.32	0.08
Perf2 → TE3	-.09	0.08	-.09	0.11	.26	0.10	.26	0.14	-.23	0.08	-.24	0.13	-.05	0.09	-.05	0.08
Perf3 → TE3	.53	0.08	.54	0.11	.21	0.09	.21	0.12	.48	0.07	.49	0.12	.48	0.07	.48	0.06
$r_{wg(j)}$ 1 → Perf1	-.05	0.15	-.06	0.15	.14	0.06	.15	0.16	.34	0.08	.34	0.14	.03	0.13	.03	0.12
$r_{wg(j)}$ 1 → Perf2	-.04	0.07	-.05	0.15	.46	0.06	.46	0.13	.07	0.11	.08	0.16	.11	0.09	.11	0.08
$r_{wg(j)}$ 1 → Perf3	-.08	0.08	-.08	0.13	-.31	0.09	-.31	0.17	-.34	0.11	-.34	0.14	-.08	0.08	-.08	0.07
$r_{wg(j)}$ 1 → TE1	-.02	0.07	-.02	0.15	.31	0.06	.31	0.13	.14	0.09	.14	0.17	.16	0.06	.16	0.05
$r_{wg(j)}$ 1 → TE3	-.10	0.05	-.10	0.10	-.13	0.08	-.13	0.14	-.03	0.09	-.03	0.13	-.03	0.06	-.03	0.05
TE1 → Perf2	.12	0.08	.12	0.15	.00	0.09	.00	0.15	.28	0.08	.28	0.15	.21	0.09	.21	0.08
TE1 → Perf3	.12	0.07	.12	0.13	.21	0.09	.21	0.15	-.08	0.09	-.08	0.16	.08	0.09	.08	0.08
TE1 → TE3	.46	0.07	.47	0.10	.67	0.07	.67	0.11	.47	0.08	.48	0.12	.50	0.08	.50	0.06

Note. AAR = After-action review; Perf 1 = performance at Time 1 (i.e., the mean of team performance scores on Test Missions 1 and 2); Eff 1 = team-efficacy at Time 1 (i.e., after Test Mission 2); $r_{wg(j)}$ 1 = within-team agreement on team-efficacy at Time 1 (i.e., after Test Mission 2); Perf 2 = performance at Time 2 (i.e., the mean of team performance scores on Test Missions 3 and 4; Perf 3 = performance at Time 3 (i.e., the mean of team performance scores on Test Missions 5 and 6; Eff 3 = team-efficacy at Time 3 (i.e. after Test Mission 6). The range of potential performance scores was from -50 to 100.

VITA

Name: Ira Anthony Schurig

Address: 4302 Bentwood Dr, College Station TX 77845

Email Address: iraschurig@yahoo.com

Education

Ph.D., Texas A&M University, College Station, Texas, 2012
Major: Industrial/Organizational Psychology

M.A., Saint Mary's University, San Antonio, Texas, 2002
Major: Industrial/Organizational Psychology

B.B.S., Hardin-Simmons University, Abilene, Texas, 1995
Major: Psychology