The comparative effectiveness of distributed and colocated team after-action reviews

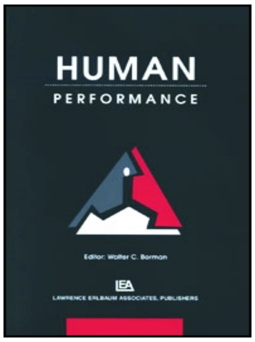
Article *in* Human Performance · October 2016

DOI: 10.1080/08959285.2016.1208662

Gonzalo J. Muñoz
Universidad Adolfo Ibáñez

Andrew Naber
RAND Corporation

Winfred Arthur, Jr.
Texas A&M University



Human Performance

ISSN: 0895-9285 (Print) 1532-7043 (Online) Journal homepage: http://www.tandfonline.com/loi/hhup20

The comparative effectiveness of distributed and colocated team after-action reviews

Steven M. Jarrett, Ryan M. Glaze, Ira Schurig, Gonzalo J. Muñoz, Andrew M. Naber, Jennifer N. McDonald, Winston Bennett Jr. & Winfred Arthur Jr.

To cite this article: Steven M. Jarrett, Ryan M. Glaze, Ira Schurig, Gonzalo J. Muñoz, Andrew M. Naber, Jennifer N. McDonald, Winston Bennett Jr. & Winfred Arthur Jr. (2016) The comparative effectiveness of distributed and colocated team after-action reviews, Human Performance, 29:5, 408-427, DOI: 10.1080/08959285.2016.1208662

To link to this article: http://dx.doi.org/10.1080/08959285.2016.1208662

Published online: 07 Oct 2016.

HUMAN PERFORMANCE
2016, VOL. 29, NO. 5, 408–427
http://dx.doi.org/10.1080/08959285.2016.1208662

The comparative effectiveness of distributed and colocated team after-action reviews

Steven M. Jarrett[a,b], Ryan M. Glaze[a], Ira Schurig[a], Gonzalo J. Muñoz[a,d], Andrew M. Naber[a], Jennifer N. McDonald[a], Winston Bennett, Jr.[c], and Winfred Arthur, Jr.[a]

[a]Texas A&M University; [b]Select International; [c]Air Force Research Lab; [d]Universidad Adolfo Ibáñez

ABSTRACT

Despite their frequent use in the military and private sectors, the *comparative* effectiveness of colocated and distributed after-action reviews (AARs) is relatively unknown. Consequently, this study examined the comparative effectiveness of colocated and distributed AARs across taskwork and teamwork outcomes. Data were obtained from 492 participants randomly assigned to 123 four-person teams who participated in one of six AAR conditions. The results indicated that teams in the AAR conditions had significantly higher performance and team efficacy scores than the teams in the non-AAR conditions. In summary, the findings highlight that regardless of the training environment or type of AAR, the AAR remains an effective method at increasing performance and other outcomes. Therefore, the use of distributed AARs does not engender the posited process losses that were hypothesized.

The after-action review (AAR; also variously referred to as the *after-event review* or *debriefing*) continues to gain increasing attention in the scholarly literature. It has been and continues to be the U.S. military's preferred method of review following collective training and/or performance in both actual and simulated settings (Meliza, Bessemer, & Hiller, 1994); the use of AARs in nonmilitary settings, such as the medical field, continues to see a dramatic increase. The U.S. Army (1993) defined 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) defined 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). In short, an AAR entails trainees systematically reviewing their performance after recently completed task or performance episodes with a focus on (a) the intended outcome; (b) the actual outcome; (c) why the intended outcome was achieved, and if not, why not; (d) setting an intended future outcome for the next performance episode; and (e) strategizing to achieve said outcome (Villado & Arthur, 2013).

Contrary to the preponderance of past research that has focused primarily on the integration of various technological advances intended to facilitate the conduct of the review (e.g., Prince, Salas, Brannick, & Orasnau, 2005), recent work has begun to empirically investigate theoretically and conceptually based design factors that may influence the effectiveness of AARs. Examples of these factors include an individual versus team focus (Schurig, Jarrett, Arthur, Glaze, & Schurig, 2011; Tannenbaum & Cerasoli, 2013), review of successful versus

CONTACT Steven M. Jarrett or Winfred Arthur, Jr. sjarrett@selectintl.com or w-arthur@tamu.edu Select International, 5700 Corporate Dr. Suite 250, Pittsburgh, PA 15206 or Department of Psychology, Texas A&M University, 4235 TAMU, College Station, TX 77843–4235.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/HHUP.

failed performance (Ellis & Davidi, 2005), and the use of objective versus subjective review systems (Villado & Arthur, 2013). Collectively, these studies have begun to provide answers to questions such as, What is the relative effectiveness of the AAR as a training method in comparison to other training methods (e.g., Arthur, Bennett, Edens, & Bell, 2003)? Is the AAR better suited for certain environments than others? Are AARs effective when conducted in a distributed training and performance environment? Answers to these questions will undoubtedly provide needed guidance to the design and implementation of AAR-based training. Hence, for instance, although Tannenbaum and Cerasoli (2013)'s meta-analysis indicated that AARs are an effective training method for improving individual and team performance, Tannenbaum and Cerasoli did not include an examination of the comparative effectiveness of distributed and colocated AARs. So, with the goal of adding to the extant empirical work examining the effectiveness of AARs, the present study examines the comparative effectiveness of distributed and colocated AAR teams on a number training outcomes, specifically, team performance, team efficacy, openness of communication, and cohesion. A second study objective is to replicate and extend the objective versus subjective AAR findings reported by Villado and Arthur (2013).

Geographic dispersion of trainees

Technological advances have facilitated and supported a continued increase in the use of distributed training. Distributed training (which in the military may take the form of distributed mission training) refers to training in which individuals interact simultaneously from different geographic locations. Distributed training is contrasted with colocated training where all team members (and the training instructor or facilitator if there is one) are situated and interact in the same physical space. An obvious advantage of distributed training is the ability to train team members who are geographically dispersed, thus reducing training costs associated with travel, lodging, and scheduling among others, without supposedly a commensurate loss in training effectiveness. Hence, the efficacy of distributed training is predicated on the premise that it is at least as effective as colocated training. Although the literature on virtual teams has indicated that distributed team members tend to demonstrate lower performance than nondistributed team members (De Guinea, Webster, & Staples, 2012; Mesmer-Magnus, DeChurch, Jimenez-Rodriguez, Wildman, & Shuffler, 2011), the focus of the present study is specifically on the effect of geographic dispersion on the effectiveness of the AAR as a training method. Said focus is also consonant with meta-analytic findings (e.g., De Guinea et al., 2012) that indicate the presence of potential moderators of the strength and direction of the virtual team findings. Thus, as previously noted, the objective of the present study is to empirically examine the comparative effectiveness of the AAR in colocated and distributed training and simulated performance environments.

A critical difference between distributed and colocated training is that with distributed training, some means of technology is required to permit communication and interaction among team members. Consequently, the type of communication technology is an important factor because it dictates the potential similarity between face-to-face and the technology-mediated interaction, and hence may influence the extent to which the two training environments, distributed and colocated, may engender similar team outcomes such as performance, efficacy, communication, and cohesion. In summary, whereas distributed training can vastly reduce the cost of training, especially in situations where training is ongoing and continuous (e.g., military teams, emergency response teams), it is important that the benefits associated with the reduction in costs are not traded off against the effectiveness of the training method. Therefore, an assessment of the comparative effectiveness of AARs in distributed and colocated settings would seem to be of some scientific and applied importance.

Technology-mediated communication in distributed training contexts

As previously noted, when teams are distributed, some means of communication, usually technology-mediated, must be used to permit interaction among team members. Clark and Brennan (1991) provided a theoretical framework for understanding the interaction between the task and the communication technology used in distributed environments by proposing a set of characteristics that differentiate face-to-face from technology-mediated communication. Specifically, they posited that interactions vary in their degree of copresence, visibility, audibility, cotemporality, simultaneity, and sequentiality—six elements that they deem critical to the potential process losses that may arise from technology-mediated interactions. Clark and Brennan defined copresence as the extent to which members are in the same location. Visibility is the extent to which members can see each other, and audibility pertains to whether team members can hear one another. Cotemporality refers to the temporal proximity between the sent and received communication, whereas simultaneity is the extent to which members can send and receive messages simultaneously. Finally, sequentiality is the extent to which messages are received in the correct order. For example, face-to-face team members speak and are heard in turn, but team members who communicate through e-mail or texting may have the order distorted if one team member misses or reads an e-mail out of sequence.

These six communication elements determine the ease with which communicators are able to understand one another's messages and form a common representation of their messages. Clark and Brennan (1991) used these six communication elements to place communication technologies on a continuum with face-to-face on one end, e-mail on the other, and audio communication falling in the middle. Clark and Brennan further posited that the ideal means of communication in a given situation is a function of the media and the purpose of the interaction.

Using Clark and Brennan (1991)'s framework, Driskell, Radtke, and Salas (2003) examined the effects of technology-mediated communication on several outcomes (e.g., cohesiveness, communication) and concluded that in general, the lack of necessary cues (e.g., audio, visual) results in a significant reduction in the achievement of these team outcomes. They concluded that "the relative loss of contextual information in computer-mediated communication can result in greater difficulty in establishing mutual knowledge" (Driskell et al., 2003, p. 317).

In terms of Clark and Brennan (1991)'s framework, synchronous audio–video communication would be the closest approximation to face-to-face communication with text-only communication (e.g., e-mail, texting) being on the furthest end of the continuum. Synchronous audio communication—the technology used as the main source of communication in the present study—falls in the middle of this continuum. We acknowledge that there are now more technologically advanced communication methods (e.g., video conferencing), however, this technology was used not only because at the time the present study was conducted, audio communication was prevalent as a means of long-distance communication and remains applicable to date, but also because of its generalizability to the situations (e.g., military action teams) that use tasks and training methods similar to that used in the present study. For instance Joint Terminal Attack Control (JTAC) controllers are forward deployed as ground-eyes on potential adversaries in areas of interest to local commanders. In a typical scenario, the JTAC identifies targets of interest for the commander who then recommends a course of action (e.g., identify the target for reconnaissance; identify and track the target for capture; and identify, track, and designate the target for action). In any of these events, the JTAC will take the direction to do what is asked, typically communicating via radio communications or chat (combat texting). Once the course of action is taken, and/or the target is neutralized, there is then a short and remotely handled AAR on the events and how well (or not well) the course of action or actions were executed.

In summary, as shown in Figure 1, in terms of Clark and Brennan (1991)'s six elements, the distributed and colocated AAR protocols used in the present study differ in copresence and visibility. Consequently, because of the relatively lower levels of media (communication) richness, including

Communication Elements	Distributed Condition	Colocated Condition
Audibility	✓	✓
Copresence	×	✓
Cotemporality	✓	✓
Sequentiality	✓	✓
Simultaneity	✓	✓
Visibility	×	✓

Figure 1. Differences between distributed and colocated conditions in terms of Clark and Brennan's (1991) six elements of technology-mediated communication.

the loss of visual (and other visually mediated nonverbal) cues associated with synchronous audio communication (Driskell et al., 2003), distributed teams were expected to display lower levels of the training outcomes.

DIstributed after-action reviews

Although limited, there is some empirical evidence supporting the effectiveness of distributed training (Dwyer, Oser, Salas, & Fowlkes, 2000; Salas, Oser, Cannon-Bowers, & Daskarolis-Kring, 2002; Townsend, Demarie, & Hendrickson, 1998). For example, Dwyer et al. (2000)'s findings indicated that event-based training in a distributed environment resulted in performance improvements across several performance episodes. However, this study did not use a control group or make any comparisons to colocated training; thus the possibility that the observed improvement in performance was merely a function of practice with the task cannot be ruled out. It was also not an AAR study.

In reference to empirical investigations of the effectiveness of distributed AARs, a detailed literature search located only two works, both unpublished doctoral dissertations (i.e., Kring, 2004; Oden, 2009) on this topic. Collectively these studies obtained fairly mixed results. For example, whereas Kring (2004) found that team-based AARs conducted in distributed environments resulted in significant improvements over baseline performance, and as would be expected, teams in the colocated condition displayed significantly higher performance than the teams in the distributed condition, Oden (2009) obtained a different pattern of results. Specifically, in an examination of the effect of different AAR formats in a distributed training environment, Oden found that teams that participated in a teleconference AAR with visual feedback (analogous to the present study's objective AAR) displayed the highest performance scores, followed by the non-AAR control condition, and then the teleconference AAR condition (analogous to the present study's subjective AAR). Thus, Oden's results indicated that in a distributed environment, non-AAR teams outperformed AAR teams. However, in the absence of a commensurate set of colocated conditions, the study does not provide insights into the *comparative* effectiveness of the AAR in distributed and colocated training environments. Absent information of this sort, it is impossible to fully evaluate the relative effectiveness of distributed and colocated AARs.

Objective versus subjective after-action reviews

A second objective of the present study was to replicate and extend the objective versus subjective AAR findings reported by Villado and Arthur (2013). As noted by Villado and Arthur (2013):

> The method of recall used during an AAR may be described as falling along a continuum, anchored by subjective on one end and objective on the other. An AAR closer to the subjective end of the continuum might

> rely exclusively on the memory of trainees … whereas an AAR closer to the objective end of the continuum might make use of recordings … to facilitate recall [and] identification of key events during the structured review. (p. 517)

Contrary to their expectations, the objective and subjective AAR teams were equally effective in terms of the specified training outcomes. Specifically, they achieved similar knowledge and performance scores, as well as similar levels of team efficacy, openness of communication, and cohesion. In a similar vein, Beaubien and Baker (2003) also failed to obtain any differences between video (i.e., objective) and nonvideo (i.e., subjective) AAR conditions. Villado and Arthur (2013) speculated that this counterintuitive finding might be attributable to the fact that both objective and subjective AAR teams "may have benefitted from both individual-level meta-cognitive and team-level macrocognitive activity brought about by the AAR" (p. 526). This reasoning is not at odds with the use of the term "reflection" or "reflexivity" in some research to describe the AAR process (e.g., see Anseel, Lievens, & Schollaert, 2009; Gurtner, Tschan, Semmer, & Nägele, 2007). That is, despite subjective AAR teams making factual errors in their recall of events associated with the performance episode, they were nevertheless able to discuss and develop strategies for improving their performance. Thus, to further investigate this unexpected result, the present study sought to replicate this finding in the context of distributed and colocated AAR teams. So, whereas we do not directly test the meta-cognition hypothesis, we examined whether the absence of differences between objective and subjective AAR teams would be observed under distributed and colocated conditions as well. Furthermore, to the extent that the benefits associated with AARs are primarily metacognitive in form, then we might even speculate that as unlikely as it may seem, a failure to obtain differences between distributed and colocated AAR teams as well might not be unexpected.

The present study

The primary objective of the present study was to examine the comparative effectiveness of distributed and colocated AARs on a number of team training outcomes, specifically team performance, team efficacy, openness of communication, and cohesion. This examination also entailed the inclusion of non-AAR distributed and colocated conditions. Consequently, on the basis of the previously discussed conceptual precepts and empirical literature, we proposed the following:

H1: Compared to non-AAR teams, both distributed and colocated AAR teams will have higher (a) team performance and will report higher levels of (b) team efficacy, (c) openness of communication, and (d) cohesion.

H2: Compared to distributed teams, colocated teams will have higher (a) team performance and will report higher levels of (b) team efficacy, (c) openness of communication, and (d) cohesion.

Whereas we also examined the effect of objective veresus subjective AARs, no formal hypotheses are posited, because consistent with Villado and Arthur (2013), we did not expect objective and subjective AAR teams to differ on any of the training outcomes regardless of their distributed or colocated status.

Method

Participants

Participants were recruited from a large southwestern public university's psychology department subject pool. The sample consisted of 492 individuals (48% female) who participated in 123 four-person teams. There were 23 teams in the colocated non-AAR condition and 20 teams each in the other conditions (see

Table 1. Demographic composition of the sample by training condition.

	Training condition													
	Colocated						Distributed							
	Non-AAR[a]		Subjective AAR[b]		Objective AAR[b]		Non-AAR[b]		Subjective AAR[b]		Objective AAR[b]		Overall	
Sex	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Female	49	53.26	42	52.50	40	50.63	40	50.00	29	36.25	34	42.50	234	47.66
Male	43	46.74	38	47.50	39[c]	49.37	40	50.00	51	63.75	46	57.50	257	52.34
No. of male participants per team[d]														
0	2	8.70	2	10.00	2	10.00	1	5.00	0	0.00	0	0.00	7	5.69
1	7	30.43	6	30.00	4	20.00	6	30.00	0	0.00	5	25.00	28	22.76
2	8	34.78	6	30.00	7	35.00	7	35.00	10	50.00	7	35.00	45	36.59
3	4	17.39	4	20.00	6	30.00	4	20.00	9	45.00	5	20.00	32	26.02
4	2	8.70	2	10.00	1	5.00	2	10.00	1	5.00	3	15.00	11	8.94
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age (in years)	18.60	0.44	18.73	0.51	18.56	0.40	18.86	0.51	19.13	1.04	19.18	0.76	18.84	1.21
Video game experience[e]	1.91	0.71	1.76	0.56	1.81	0.70	1.76	0.56	1.83	0.66	1.75	0.61	1.81	0.65

Note. AAR = after-action review.
[a]*n* = 23 teams. [b]*n* = 20 teams. [c]Demographic data for one participant in the colocated objective AAR training condition were missing. [d]*n* = number of teams (*N* = 123 four-person teams). [e]Video game experience was measured using a 3-point scale: 1 (*novice*), 2 (*average*), and 3 (*expert*).

Table 1). Participants reported a mean age of 18.84 years ($SD = 0.70$) 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 of 1 (*novice*), 2 (*average*), and 3 (*expert*). Furthermore, the six conditions did not differ in terms of their video game experience scores, $F(5, 486) = 1.22$, $p > .05$, $\eta^2 = .01$. In addition to course credit for their participation, 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 study condition were awarded \$80, \$40, and \$20, respectively. Overall and condition-specific demographic information for the study participants are presented in Table 1.

Design

The study design was a 2 (geographic dispersion: distributed vs. colocated) × 3 (type of AAR review: non-AAR vs. objective AAR vs. subjective AAR) × 3 (session: Sessions 1–3) repeated measures experimental design. Geographic dispersion and type of AAR review served as the between-subjects independent variables, and session served as the repeated or within-subjects independent variable. Four dependent variables were measured at various times throughout the study protocol to assess skill, efficacy, openness of communication, and cohesion in teams. Table 2 presents an overview and summary of the study protocol.

Measures

Performance task: Steel Beasts Pro PE

Steel Beasts Pro PE version 2.370 (eSim Games, 2007) was used as the team training and performance task. Steel Beasts is a cognitively complex, PC-based battle tank simulation allowing multiple players to complete a simulated mission against enemy tanks. (In the present study, the enemy tanks were operated by the computer.) It is an ecologically valid analogue of the types of tasks (e.g., psychomotor, cognitive, information processing, and team coordination) that are

Table 2. Schedule of activities for each session by training condition.

Session	Scheduled activities					
0	Informed consent Video game experience Demographics Participants randomly assigned to teams Team members randomly assigned to roles Teams randomly assigned to training condition Individual tutorials					
	Training conditions					
	Colocated			Distributed		
	Non-AAR[a]	Subjective AAR[b]	Objective AAR[b]	Non-AAR[b]	Subjective AAR[b]	Objective AAR[b]
1	Planning Test Mission 1A Planning Test Mission 1B Filler Task Team efficacy 1	Planning Test Mission 1A Planning Test Mission 1B AAR Team efficacy 1	Planning Test Mission 1A Planning Test Mission 1B AAR Team efficacy 1	Planning Test Mission 1A Planning Test Mission 1B Filler Task Team efficacy 1	Planning Test Mission 1A Planning Test Mission 1B AAR Team efficacy 1	Planning Test Mission 1A Planning Test Mission 1B AAR Team efficacy 1
2	Planning Practice Mission 1 Filler Task Planning Test Mission 2A Planning Test Mission 2B Filler Task	Planning Practice Mission 1 AAR Planning Test Mission 2A Planning Test Mission 2B AAR	Planning Practice Mission 1 AAR Planning Test Mission 2A Planning Test Mission 2B AAR	Planning Practice Mission 1 Filler Task Planning Test Mission 2A Planning Test Mission 2B Filler Task	Planning Practice Mission 1 AAR Planning Test Mission 2A Planning Test Mission 2B AAR	Planning Practice Mission 1 AAR Planning Test Mission 2A Planning Test Mission 2B AAR
3	Planning Practice Mission 2 Filler Task Planning Test Mission 3A Planning Test Mission 3B Team efficacy 2 Openness of comm. Cohesion	Planning Practice Mission 2 AAR Planning Test Mission 3A Planning Test Mission 3B Team efficacy 2 Openness of comm. Cohesion	Planning Practice Mission 2 AAR Planning Test Mission 3A Planning Test Mission 3B Team efficacy 2 Openness of comm. Cohesion	Planning Practice Mission 2 Filler Task Planning Test Mission 3A Planning Test Mission 3B Team efficacy 2 Openness of comm. Cohesion	Planning Practice Mission 2 AAR Planning Test Mission 3A Planning Test Mission 3B Team efficacy 2 Openness of comm. Cohesion	Planning Practice Mission 2 AAR Planning Test Mission 3A Planning Test Mission 3B Team efficacy 2 Openness of comm. Cohesion

Note. 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. AAR = after-action review; comm. = communication.
[a]n = 23 teams. [b]n = 20 teams.

trained in operational settings as reflected in the fact that it is used by several defense and military agencies (e.g., the armies of Denmark, the Netherlands, Spain, and Sweden, among others). The simulator uses accurate replicas of U.S. (participants') and Russian tanks (computer-operated enemy) to simulate an armored warfare battle. Each user-controlled tank required two participants, one serving as the gunner and the other as the commander/driver. Hence, a team consisted of a two-tank platoon with each tank consisting of two participants. The teams performed the task via four-networked computers, with each member of the four-person team at their own computer. Team members communicated with each other using networked-linked voice-activated microphones and headsets.

Multiple first-person perspective views were available to each participant, depending on their 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. Performance missions were 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.

Steel Beasts Pro PE missions. The study protocol consisted of three sessions with two test missions in each of the three sessions. The same mission map (see Figure 2) was used for all six test missions and the two practice missions. Each test mission required a team to destroy 10 enemy tanks while they were en route to a target destination. A mission ended when (a) the team completed all mission objectives, (b) all participant tanks were destroyed, or (c) the 10-min time limit expired. The two practice missions were identical to test missions with the exception that for practice missions teams were allowed 15 min, which they could use for either planning or interacting with the simulator. In addition, the first practice mission provided participants with suggested waypoints for optimal performance of the missions (see Figure 2). Participants were also informed that their practice mission scores would not count toward their performance scores.

Performance scores were obtained at the team level. Teams earned points for the number of enemy tanks destroyed (5 points per tank) and advancing beyond certain boundaries (2.5 points per tank per boundary crossed [e.g., Sierra and Alpha in Figure 2] and 12.5 for each team tank that reached the objective). Teams lost points for destroying one of their own tanks (fratricide; −50 points). Thus, the total possible points ranged from −50 to +100. 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 each mission.

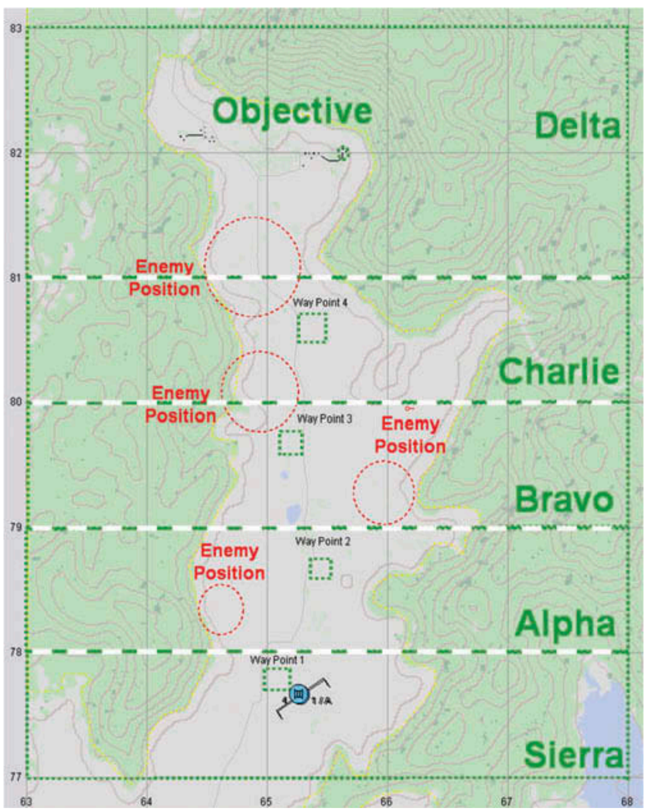


Figure 2. Mission map. *Note*. Waypoints were displayed for only the first practice mission.

Team efficacy

A modified version of Arthur, Bell, and Edwards (2007)'s 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 Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). Team efficacy scores were calculated using the average 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).

Openness of communication, and cohesion

Openness of communication and cohesion were assessed using a 10-item measure that consisted of four communication and six cohesion items. The 10 items were selected from Barry and Stewart (1997)'s Group Process measure, and modified to fit the present performance task. Participants provided their ratings using a 5-point Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). The internal consistency estimate for openness of communication scores at the individual-level of analysis (N = 492) was .72 and was .87 for cohesion scores. Openness of communication and cohesion scores were calculated using the average of the mean individual-level item responses for each subscale.

Demographics

Participants reported their age, sex, experience with video games, and whether they had any 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., 1 = *novice*, 2 = *average*, 3 = *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.

Training manipulation

During the study, participants operated the simulator as a team to complete both the test and practice missions. All teams completed the same test missions, regardless of training condition. The events that followed each team performance mission depended on the training condition to which the team was assigned.

Geographic dispersion

In contrast to the colocated training condition, participants on distributed teams were geographically separated. Specifically, the two participants assigned to one tank remained together in the primary laboratory, and the individuals in the other tank were escorted to a second laboratory (henceforth referred to as the distributed laboratory) in a separate building on campus. Although there were two team members in each laboratory, all communication among the four team members during the performance events *and* the AARs was conducted via voice-activated microphones and headphones. The distributed laboratory had the same equipment as the primary laboratory. That is, participants in the distributed laboratory operated Steel Beasts using the same make and model of monitors, keyboards, mice, and joysticks as their teammates in the primary laboratory and communicated with each other and their teammates in the primary laboratory via voice-activated microphones and headphones. All four members of colocated teams were situated in the primary laboratory. The research protocol entailed the running of only one team at a time.

Non-AAR training conditions

Once a team mission ended, participants assigned to the non-AAR conditions (both colocated and distributed) completed a filler task that was unrelated to Steel Beasts.[1] This was to maintain the same time interval between activities for the non-AAR and AAR training conditions.

Subjective AAR training conditions

After the two practice missions and test missions 1B and 2B (see Table 2), participants in the subjective AAR condition (both colocated and distributed) participated in a 10-min AAR, monitored by a facilitator. Prior to the first AAR, the facilitator explained the AAR process to team members and provided teams with a form that detailed each step of the AAR process. After introducing participants to the AAR process, the facilitator intervened during AARs only to ensure that teams completed each step of the AAR in order and within the specified time limits.

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. Participants then completed the specified paper-and-pencil measures as per Table 2.

For the distributed condition, each AAR was guided by a facilitator who was located in the primary laboratory. Each AAR was conducted in the same manner as the colocated subjective AARs except that participants communicated via voice-activated microphones and headphones instead of the face-to-face manner described for the colocated subjective AAR condition. However, similar to the colocated condition, once they had completed the AAR, participants next completed any measures scheduled to follow the AAR session, and if there were no measures, then they immediately logged back into the simulator.

Objective AAR training conditions

Participants in the objective AAR training condition (both colocated and distributed) also participated in a 10-min AAR (monitored by a facilitator) after the two practice missions and test missions 1B and 2B (see Table 2). However, unlike the subjective conditions, participants in the objective AAR conditions reviewed the progress of their most recently completed mission using the simulator's review tool, operated by the facilitator. By means of a display that was projected on a large screen, 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 tank's perspective, the enemy's perspective, or a God's-eye view of the mission). Otherwise, objective AAR teams followed the same review procedures as subjective AAR teams, and after the AAR they also completed the specified paper-and-pencil measures as per Table 2.

For the distributed objective AAR condition, each AAR was conducted in the same manner as the colocated objective AARs, except that participants communicated via voice-activated microphones and headphones rather than face-to-face. In addition, the primary laboratory had a projector screen and the distributed laboratory had a large monitor that provided the same video playback simultaneously to all four participants (two in the primary laboratory and two in the distributed lab). Similar to the colocated condition, once they had completed the AAR, participants then completed any measures scheduled to follow that mission, or if there were no measures, then they immediately logged back into the simulator.

Procedure

The study protocol lasted 5 hr and was divided into three phases. Participants were first randomly assigned to not only teams but also specific roles (i.e., gunner or commander/driver) within the team, as well as a specific tank. Teams were then also randomly assigned to conditions. Prior to the commencement of any of the study activities, individuals in the distributed condition were separated by tank with two individuals in the primary laboratory (one gunner and one commander/driver) and two individuals (also one gunner and one commander/driver) in the distributed laboratory with a proctor at each location. Steps were taken to minimize the amount of communication between the trainees prior to the start of the experiment. For the colocated condition, all four participants (i.e., both tanks) were situated in the primary laboratory.

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 nine 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 used the same procedure. Six of the training tutorials focused on tasks relevant to a participant's role, and the remaining three tutorials focused on tasks relevant to their teammate's role.

Upon completing the tutorials, participants then moved to the team training phase. This phase of the protocol commenced with participants being shown how to use the voice-activated microphones and headphones. At this point, all participants put on their headsets, which were used as the main form of communication for team interactions with the exception of colocated AAR sessions. Participants were asked to demonstrate their ability to use the microphones and headsets. Once all participants (distributed and colocated) had their headsets on and they were working properly, they were given verbal instructions by the proctor in the primary laboratory to begin the first team mission.

Each team mission began with a planning period. Participants were allowed 2 min to review the mission briefing and map, formulate a strategy, and discuss the strategy with their teammates during the planning period. Teams were allowed to begin the mission prior to the 2-min time limit if all team members were ready to do so and agreed to it. Otherwise, the team mission began after 2 min had expired. Teams were allowed 10 min to complete each team mission. The simulator displayed the mission runtime. Trainees completed the test missions in two mission blocks (e.g., Mission 1A and Mission 1B; see Table 2). The first two test missions were followed by a practice session that lasted 15 min. The practice session scores were not counted toward the teams' performance score. In addition, if a practice mission ended early, participants were able to restart the mission and use the entire 15 min. Once a team completed a particular mission or the mission was terminated, teams in the AAR conditions participated in the AAR process. Team mission briefing (2 min), team mission (10 min), and AAR (10 min) time limits were established and deemed to be sufficient on the basis of pilot testing.

Results

The individual-level data for team efficacy, openness of communication, and cohesion were evaluated to justify aggregation to the team level. As shown in Table 3, agreement and reliability indices (i.e., $r_{wg[1]}$, $r_{wg[j]}$, r^*_{wg}, ICC_1, and ICC_2; James, Demaree, & Wolf, 1984; Lindell, Brandt, & Whitney, 1999) indicated that it was appropriate to do so. Team-level scores for these variables were

Table 3. Individual-level team efficacy, openness of communication, and cohesion median agreement indices and intraclass correlation coefficients (ICCs).

Measures	$R_{wg(1)}$	$R_{wg(j)}$	R^*_{wg}	ICC_1	ICC_2
Team efficacy (6 items)					
Time 1 (Session 1)	.80	.91	.65	.13	.52
Time 2 (Session 3)	.92	.92	.68	.15	.55
Openness of communication (4 items)	.88	.93	.75	.17	.50
Cohesion (6 items)	.92	.89	.66	.12	.56

Note. $N = 492$; $k = 123$. Median $r_{wg(1)}$ and $r_{wg(j)}$ were calculated using the formulas presented by James et al. (1984). Median r^*_{wg} was calculated using the formulas presented by Lindell et al. (1999).

Table 4. Means, standard deviations, and intercorrelations of study variables.

Variable	*M*	*SD*	1	2	3	4	5	6
1. Mission 1	27.92	6.33	—					
2. Mission 2	35.62	9.11	.24*	—				
3. Mission 3	39.33	10.56	.09	.32*	—			
4. Team efficacy 1	3.31	0.47	.13	.25*	.15*	—		
5. Team efficacy 2	3.54	0.52	−.01	.20*	.52*	.54*	—	
6. Openness of comm.	4.11	0.34	.09	.13	.34*	.31*	.49*	—
7. Cohesion	3.99	0.39	.07	.09	.33*	.33*	.56*	.83*

Note. N = 123 teams. comm. = communication.
*$p < .05$ (one-tailed).

Table 5. Means and standard deviations of study variables by training condition.

	Training condition											
	Colocated						Distributed					
	Non-AAR[a]		Subjective AAR[b]		Objective AAR[b]		Non-AAR[b]		Subjective AAR[b]		Objective AAR[b]	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Mission 1	28.32	5.09	28.19	6.92	28.38	4.37	29.19	6.57	25.69	6.79	28.00	7.91
Mission 2	31.74	9.96	35.81	9.05	36.56	9.61	36.25	8.23	35.88	9.06	38.06	8.27
Mission 3	33.53	9.93	41.38	9.36	41.25	13.78	39.06	9.34	40.56	10.47	41.06	8.65
Team efficacy 1	3.06	0.45	3.44	0.33	3.40	0.47	3.25	0.27	3.49	0.58	3.29	0.51
Team efficacy 2	3.10	0.53	3.65	0.34	3.61	0.61	3.52	0.46	3.79	0.44	3.67	0.41
Openness of comm.	3.80	0.30	4.24	0.28	4.23	0.25	4.04	0.43	4.30	0.26	4.08	0.25
Cohesion	3.63	0.37	4.12	0.32	4.08	0.28	3.93	0.44	4.20	0.34	4.06	0.32

Note. All scales are on a 5-point Likert scale, except performance, which can range from −50 to 100. AAR = after-action review; comm. = communication.
[a]*n* = 23 teams. [b]*n* = 20 teams.

accordingly created by averaging the individual-level scores within teams. Team performance scores were obtained at the team level, and so there was no need to aggregate these scores. Team-level descriptive statistics and intercorrelations for the study variables are presented in Table 4 and Table 5, respectively.

H1a to H1d

Two separate mixed analyses of variance (ANOVAs) were run for team performance and team efficacy.[2] Training condition (non-AAR vs. AAR [collapsed across geographic dispersion and review type]) served as the independent variable, and session (Missions 1–3 for team performance, and team efficacy Time 1 and 2) served as the within-subjects independent variable. Independent *t*-tests were run for openness of communication and cohesion.

H1a: Team performance

H1a posited that AAR teams will have higher performance scores than non-AAR teams. Using a 2 × 3 mixed ANOVA, the between-subject main effect was statistically significant, $F(1, 121) = 3.55$, $p < .05$, $\eta^2 = .03$, indicating that the AAR teams obtained higher performance scores than the non-AAR teams. In addition, the within-subjects analysis indicated that teams improved across sessions, $F(2, 242) = 53.76$, $p < .05$, $\eta^2 = .44$, and there was a statistically significant Training Condition × Session interaction, $F(2, 242) = 4.61$, $p < .05$, $\eta^2 = .04$, demonstrating that the level of performance improvement depended on the training condition (i.e., AAR vs. non-AAR). Thus, H1a was supported.

Additional analyses were conducted to examine whether the performance of non-AAR and AAR teams differed as a function of geographic dispersion. The results of these analyses indicated a significant Training Condition × Geographic Dispersion interaction such that, whereas there was no difference between the distributed non-AAR teams and the distributed AAR teams, $F(1, 58) = 0.46$, $p > .05$, $\eta^2 = .01$, colocated AAR teams obtained higher performance scores than colocated non-AAR teams, $F(1, 61) = 7.01$, $p < .05$, $\eta^2 = .11$.

H1b: Team efficacy

H1b, which had posited that AAR teams will report higher levels of team efficacy than non-AAR teams, was supported, $F(1, 121) = 17.03$, $p < .05$, $\eta^2 = .14$. Furthermore, team efficacy scores for both AAR and non-AAR teams increased across the two efficacy administrations, $F(1, 121) = 22.07$, $p < .05$, $\eta^2 = .18$, but the Training Condition × Session interaction was not significant, $F(1, 121) = 1.66$, $p > .05$, $\eta^2 = .02$.

Additional analyses were conducted to examine whether the levels of efficacy reported by non-AAR and AAR teams would differ as a function of geographic dispersion. These analyses were limited to only the Time 2 efficacy scores, as it was the more distal of the two. The results of these analyses indicated that colocated AAR teams reported higher levels of efficacy than colocated non-AAR teams ($d = 1.04$, $p < .05$; see Table 6); the same pattern of results were obtained for distributed teams as well, but they were not statistically ($d = 0.48$, $p > .05$; see Table 6).

H1c and H1d: Openness of communication and cohesion

Two separate between-groups *t*-tests were performed to test H1c and H1d, which had posited that AAR teams would report higher levels of openness of communication and cohesion than non-AAR teams. The results supported H1c (communication), $t(121) = 5.13$, $p < .05$, $d = 0.92$. In addition, the difference between colocated AAR and colocated non-AAR teams ($d = 1.50$, $p < .05$; see Table 6) was larger than the difference between distributed AAR and distributed non-AAR teams ($d = 0.52$, $p < .05$; see Table 6).

A similar pattern of results was obtained for cohesion in that AAR teams reported higher levels of cohesion than non-AAR teams, $t(121) = 5.25$, $p < .05$, $d = 0.94$. Hence, H1d was supported. In addition, the difference between colocated AAR and colocated non-AAR teams ($d = 1.40$, $p < .05$; see Table 6) was larger than that observed between distributed AAR and distributed non-AAR teams ($d = 0.62$, $p < .05$; see Table 6).

H2a to H2d

The analyses used to test H1a to H1d were used to test H2a to H2d as well. However, in this instance the distributed versus colocated conditions served as the between-subjects independent variable.

Table 6. Pairwise standardized mean differences (ds) by training condition.

	Performance					
Comparison	Mission 1	Mission 2	Mission 3	Team efficacy[a]	Openness of comm.	Cohesion
AAR vs. non-AAR	−0.20	0.29	0.48*	0.74*	0.92*	0.94*
Colo.	−0.01	0.46*	0.72*	1.04*	1.50*	1.40*
Dist.	−0.35	0.09	0.19	0.48	0.52*	0.62*
Colo. AAR vs. dist. AAR	0.24	−0.09	0.05	−0.22	−0.19	−0.09
Obj. AAR vs. subj. AAR	0.21	0.14	0.02	−0.18	−0.42	−0.32
Colo.	0.03	0.08	−0.01	−0.08	−0.03	−0.13
Dist.	0.35	0.25	0.05	−0.28	−0.57*	−0.42

*Note. D*s were computed by subtracting the second condition from the first such that a positive *d* indicates the teams in the first condition had a higher score on each variable than the second condition. comm. = communication; AAR = after-action review; colo. = colocated; dist. = distributed. Obj. = objective; Subj. = subjective.

[a]Time 2 team efficacy.

*$p < .05$ (one-tailed).

H2a: Team performance

H2a, which had posited that colocated AAR teams will have higher performance scores than distributed AAR teams was not supported, $F(1, 78) = 0.11$, $p > .05$, $\eta^2 = 0.00$. In addition, whereas there was significant improvement across the performance sessions, $F(2, 156) = 57.62$, $p < .05$, $\eta^2 = .42$, the Training Condition × Session interaction was not statistically significant, $F(2, 156) = 0.42$, $p > .05$, $\eta^2 = 0.00$, indicating that performance improvement did not depend on the training condition.

Additional analyses were conducted to examine whether the performance of distributed and colocated teams differed as a function of the type of review (i.e., objective vs. subjective). The results of these analyses indicated the Geographic Dispersion × Type-of-Review interaction was not significant, $F(2, 77) = 1.33$, $p > .05$, $\eta^2 = .03$; there were no mean differences between the colocated subjective and objective AAR teams, $F(1, 38) = 0.02$, $p > .05$, $\eta^2 = 0.00$, or the distributed subjective and objective AAR teams, $F(1, 38) = 0.86$, $p > .05$, $\eta^2 = 0.00$.

H2b: Team efficacy

H2b, which had posited that colocated AAR teams will report higher levels of team efficacy than distributed AAR teams, was not supported, $F(1, 78) = 0.93$, $p > .05$, $\eta^2 = .01$. Furthermore, whereas team efficacy scores for both distributed and colocated AAR teams increased across the two efficacy administrations, $F(1, 78) = 25.79$, $p < .05$, $\eta^2 = .33$, the Geographic Dispersion Condition × Session interaction was not significant, $F(1, 78) = 1.44$, $p > .05$, $\eta^2 = .02$.

Additional analyses were conducted to examine whether the levels of efficacy reported by distributed and colocated AAR teams would differ as a function of review type. The results of these analyses indicated that, whereas distributed objective AAR teams reported lower levels of efficacy than their subjective counterparts ($d = -0.28$, $p > .05$; see Table 6), there was a smaller difference between colocated objective and subjective teams ($d = -0.08$, $p > .05$; see Table 6). However, none of the above effects were statistically significant.

H2c and H2d: Openness of communication and cohesion

Two separate between-groups *t*-tests were performed to test H2c and H2d, which had posited that colocated AAR teams would report higher levels of openness of communication and cohesion than distributed AAR teams. The results failed to support H2c, $t(78) = 0.86$, $p > .05$, $d = -0.19$. In addition, there was no difference in the levels of reported openness of communication between colocated objective and subjective AAR teams ($d = -0.03$, $p > .05$; see Table 6). However, there was a statistically significant difference between distributed objective and subjective AAR teams ($d = -0.57$, $p < .05$; see Table 6) on openness of communication.

A similar pattern of results was obtained for cohesion in that again there was no difference in the levels of cohesion reported by colocated and distributed teams, $t(78) = 0.46$, $p > .05$, $d = -0.09$. Hence, H2d was not supported. In addition, there were no differences in the levels of cohesion reported by colocated objective and subjective AAR teams ($d = -0.13$, $p > .05$; see Table 6) and distributed objective and subjective AAR teams ($d = -0.42$, $p > .05$; see Table 6).

Objective versus subjective AARs

In a replication of Villado and Arthur (2013), the results were consistently indicative of the absence of a difference between objective and subjective AAR teams on all the dependent variables of interest: performance, $F(1, 76) = 0.58$, $p > .05$, $\eta^2 = 0.00$; team efficacy, $F(1, 76) = 1.26$, $p > .05$, $\eta^2 = .02$; cohesion, $F(1, 76) = 1.80$, $p > .05$, $\eta^2 = .02$). The one exception was for communication, in which subjective AAR teams reported higher levels of openness of communication than objective AAR teams, $F(1, 76) = 3.66$, $p < .05$, $\eta^2 = .04$. The general absence of a difference between objective and subjective AAR teams was maintained across the colocated and distributed conditions as well: performance, $F(1, 76) = 0.31$, $p > .05$, $\eta^2 = 0.00$; team efficacy, $F(1, 76) = 0.45$, $p > .05$, $\eta^2 = .01$;

cohesion, $F(1, 76) = 0.48$, $p > .05$, $\eta^2 = .01$. However, once again, the only exception was the finding that distributed subjective AAR teams reported higher levels of openness of communication than the corresponding objective teams, $F(1, 76) = 3.45$, $p < .05$, $\eta^2 = .04$.

Discussion

An objective of present study was to examine the comparative effectiveness of distributed and colocated AAR teams in terms of performance, team efficacy, openness of communication, and cohesion. The use of a variety of outcome variables is in recognition of the principle that the effectiveness of a specified training method may vary as a function of the outcome of interest (Arthur, Stanush, & McNelly, 1998; Schmidt & Björk, 1992). A second objective was to further examine the effect of review type (i.e., objective vs. subjective reviews). To summarize the results, first, they generally indicated that AAR teams had higher levels of performance and reported higher levels of team efficacy, openness of communication, and cohesion than non-AAR teams. Furthermore, the general pattern of results indicated that the differences between colocated AAR and colocated non-AAR teams on the dependent variables were larger than those observed for the differences between distributed AAR and distributed non-AAR teams.

Second, in the examination of the effect of geographic dispersion, there were no differences between colocated and distributed AAR teams on the dependent variables of interest. Furthermore, with the exception of openness of communication, for which distributed subjective AAR teams reported higher levels than objective teams, the absence of a difference between objective and subjective AAR teams found by Villado and Arthur (2013) was replicated—a finding that is also consistent with that obtained by Savoldelli et al. (2006), who failed to obtain a difference between oral and video-assisted oral feedback. A set of post hoc power analyses indicated that the lack of statistically significant findings for the between-subjects tests (e.g., objective vs. subjective comparison power = .90) were unlikely a result of the sample size.

General effectiveness of AARs

Consonant with its implementation in the present study, AARs are commonly considered to be a taskwork method that is focused on improving task-related knowledge and/or performance (Kozlowski & Bell, 2003). However, the results of the present study, in conjunction with those of Villado and Arthur (2013), demonstrate that AARs can also influence teamwork outcomes such as cohesion and openness of communication. This is noteworthy because both taskwork and teamwork have been demonstrated to be necessary for effective team performance (Salas, Bowers, & Rhodenizer, 1998). Hence, although unintended, it would seem that taskwork-focused AARs by virtue of the review session also provide the opportunity for the development of teamwork (e.g., communication skills, team cohesion).

Colocated versus distributed AARs and objective versus subjective reviews

Baltes, Dickson, Sherman, Bauer, and LaGanke's (2002) meta-analysis, albeit of decision-making teams, concluded that technology-mediated teams (i.e., distributed teams) are rarely more effective than colocated teams. This conclusion is consonant with propositions originating from media richness theories (e.g., Clark & Brennan, 1991) that would suggest that because of the relatively lower levels of media (communication) richness, distributed teams would be expected to display lower levels on the training outcomes of interest. Surprisingly, contrary to the preceding, the present study failed to obtain differences between colocated and distributed AAR teams. There are several plausible explanations for what at first appears to be contrarian finding but may not really be.

First, the absence of a difference may be due to the instructional design properties of the AAR. Specifically, Sitzmann, Kraiger, Stewart, and Wisher's (2006) meta-analysis demonstrated that the instructional design of training methods moderates the relationship between colocated and distributed performance. For example, Sitzmann et al. found that when the training method incorporated learner control and feedback, there was no performance difference between colocated and distributed training. However, there were significant performance differences between the colocated and distributed training conditions when the training methods did not incorporate these instructional design elements. Thus, the structure of the AAR (i.e., the presence of instructional design characteristics such as learner control and feedback) may be a critical design feature that mitigates performance differences between colocated and distributed AAR teams.

A second explanation is the metacognitive explanation forwarded by Villado and Arthur (2013). Specifically, as previously alluded to, to the extent that the benefits associated with AARs are primarily metacognitive in form, it may indeed not be that unexpected that there are no differences between colocated and distributed AAR teams. This explanation would also account for the absence of a difference between objective and subjective AARs as well. That is, it would seem that in spite of inaccuracies in the recall of performance events during the review session, there may be some metacognitive benefits to simply participating in the AAR regardless of the extent to which the performance review is accurate (Ellis, Mendel, & Aloni-Zohar, 2009) or even related to the team's *own* performance (Ellis, Ganzach, Castle, & Sekely, 2010). Metacognition consists of two components: knowledge acquisition—accumulated knowledge about cognitive process—and regulation of knowledge, which is the use of regulatory strategies to facilitate cognitive performance (Baker & Brown, 1984; Flavell, 1979). Thus, the AAR may represent a type of knowledge regulation and subsequently influence learning and performance by facilitating the cognitive process, regardless of the geographic dispersion or type of AAR. Clearly, this metacognitive explanation warrants direct research attention.

Implications

The major implications of the present work arise from the failure to obtain differences between colocated and distributed, and objective and subjective AAR teams. Concerning the former, the results suggest that the use of distributed AAR teams does not necessarily engender performance and process losses that diminish the effectiveness of AARs. Thus, the use of AARs to train geographically dispersed individuals and teams may be a viable option. However, it is worth noting that although not a central focus of the present study, we collected trainee reaction data to evaluate participants' reactions to the colocated and distributed conditions. The results of these supplementary data indicated that participants in the colocated AAR condition felt more favorably[3] toward it (M = 3.72, SD = 0.42) than those in the distributed condition (M = 3.36, SD = 0.36, d = 0.92, $p < .05$). Hence, to the extent that training reactions are an important concern, this difference may be consequential.

The second implication arises from the absence of a difference between objective and subjective AAR teams. As also noted by Villado and Arthur (2013), this suggests that AARs may be amenable to a fairly wide range of tasks and performance environments, particularly in those where it may not be practical or feasible to record or otherwise document performance for subsequent objective review.

Limitations and directions for future research

There are some characteristics of the present study that serve as avenues for future research. First, the use of only audio communication (among the trainees) in the distributed condition may have disadvantaged that condition; that the use of video and audio communication in the distributed condition may have been a fairer comparison to the colocated condition. However, if this was indeed disadvantageous, then it should have contributed to more favorable outcomes for the colocated condition, which was not the case. That being said, this is still an issue worth examining. An additional potential limitation pertains to the form of the geographic dispersion of team members in the distributed condition. Specifically, instead of having all four team members separated from each other, in the present study they were separated in pairs with each pair (representing a tank) in a common location. However, participants in a pair did not have a direct line of sight to one another, and all communication during the performance events *and* AARs was completed through the voice-activated microphones and headphones. Furthermore, in the present study *both* the performance of the team task and the AAR were distributed (as previously described) or colocated. However, one could conceivably have a situation in which performance is distributed but the subsequent AAR is colocated, or vice versa. Hence, examinations of the effect of having *all* team members separated from each other, and/or one activity (e.g., performance) being distributed and the other (e.g., the AAR) being colocated, would be elucidative. A related line of future research might be the extent to which the distributeness of team members facilitates or hinders the effectiveness of AARs as a means of facilitating the development of teamwork skills where, unlike the present study, that is the primary focus of the AAR.

Second, the use of an action task serves as a potentially important boundary condition. For instance, although it was not an AAR study, Baltes et al.'s (2002) meta-analysis of decision-making tasks found colocated teams to achieve superior outcomes than distributed teams. Decision-making tasks are characterized as tasks that require reaching consensus on issues with no right answer and/or that have high levels of uncertainty and complexity and are nonroutine (Sundstrom, De Meuse, & Futrell, 1990). On the other hand, action tasks have discernable action steps and activities that are implemented by the team. Hence, it would seem that in contrast to decision-making tasks, action tasks by their very nature may be less susceptible to the process losses associated with team member dispersion. So, because most recent AAR research has used action-based tasks, additional research investigating the role of task type as a boundary condition is likely to be informative.

Third, another important boundary condition is that the present study's action task was completed by all participants in the same, albeit digital, environment. Consequently, they all saw the same environment and experienced the same events. However, to the extent that the AAR is conducted with teams performing their work or tasks somewhat separately—for example, a team that compiles a software program by independently creating portions of code and then working together to put them into a cohesive whole—one may obtain effects that are different than those reported here. In short, again, the nature of the task may be an important boundary condition.

Fourth, it has been observed by Ackerman and Cianciolo (2000) that massed protocols such as that used in the present study (5 hr) do not provide teams performing complex tasks adequate time to develop task proficiency or learn to work as a team. So, for instance, as relative novices, trainees may be overwhelmed by the amount of information presented in the objective AAR; in contrast, as team members become experts they may be more capable of processing objective information specific to their past performance episode(s). Thus, the massed protocol may engender a situation in which the subjective and objective AARs are functionally similar and trainees are incapable of garnering the benefits of the objective AAR. Hence, examinations of whether teams at varying levels of the skill acquisition (i.e., experts vs. novices) are capable of utilizing the AAR information differently may be informative.

Related to the massed nature of the protocol, the spacing of the AAR or proximity of the AAR to the performance episode represents a potentially important factor in examining the efficacy of objective AARs (Williams & Watson, 2004). The present study had the teams perform the AAR immediately after the performance episode, so teams were more likely to have the details of the previous performance episode accessible. However, in situations where there may be long gaps between the performance episode and the AAR, the use of objective AARs may be more beneficial because they do not rely on the teams' ability to recall critical incidents from the performance episode. The longer the time interval between the performance episode and the review, the lower the likelihood that teams will be able to accurately assess their performance without the help of memory aids (e.g., recordings or diaries). Thus, future research could examine how the temporal proximity of the AAR to the performance episode affects the comparative effectiveness of objective and subjective AARs.

Conclusion

The present study sought to examine the comparative effectiveness of colocated and distributed AARs. As with the results obtained for objective and subjective reviews (see also Savoldelli et al., 2006; Villado & Arthur, 2013), and the results indicated that colocated and distributed AAR teams did not differ in terms of both taskwork (i.e., performance) and teamwork outcomes (i.e., efficacy, cohesion, and openness of communication). These findings suggest that AARs may be amenable to a fairly wide range of task and performance settings (e.g., see Yorio & Wachter, 2014). Finally, this study also contributes to the growing body of empirical literature evaluating and documenting the factors and characteristics that serve as boundary conditions for the effectiveness of the AAR as a training method. Nevertheless, although there is increasing burgeoning empirical evidence to support the assertion that AARs are an effective training method, there is a need for research that explores and clarifies the underlying dynamics of why AARs work.

Funding

This research was sponsored by contracts and grants from the U.S. Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Research Division, Mesa, AZ, awarded to Winfred Arthur, Jr. The views expressed herein are those of the authors and do not necessarily reflect the official position or opinion of the sponsors or their respective organizations.

Notes

1. The filler task was automobile driving related. Specifically, participants provided importance ratings and mental model similarity ratings of 12 driving-related concepts (e.g., braking, yellow traffic lights, turn signal). Participants spent the same amount of time on the filler task as the AAR session, and there was no reason to expect the filler task to improve or decrease performance on Steel Beasts. In addition, because the task was completed independently, there was no reason to expect that the subject or method of the task impacted the team metrics that were of interest in the current study.
2. Consistent with previous work (e.g., Jarrett et al., 2010; Sanchez-Ku & Arthur, 2000) indicative of sex differences on action-based tasks with high psychomotor demands such as that used in the present study, the sex composition of teams was related to team performance and efficacy. Consequently, because sex composition was not experimentally manipulated, the performance and team efficacy analyses were ran with sex composition as a covariate. Given the similarity of the results to the analysis of covariance to those of the ANOVAs, to maintain consistency with the other dependent variables, only the results of the ANOVAs are reported here. The analysis of covariance results are available from the first author upon request.
3. Affective training reactions were measured using six items ($\alpha = .89$) using a 5-point Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). An example item is "I enjoy participating in after-action reviews."

References

Ackerman, P. L., & Cianciolo, A. T. (2000). Cognitive, perceptual-speed, and psychomotor determinants of individual differences in skill acquisition. *Journal of Experimental Psychology: Applied*, *6*, 259–290.

Anseel, F., Lievens, F., & Schollaert, E. (2009). Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes*, *110*, 23–35.

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 Jr., W., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology*, *88*, 234–245.

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

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson, R. Barr, M. Kamil & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 353–394). New York, NY: Longman.

Baltes, B. B., Dickson, M. W., Sherman, M. P., Bauer, C. C., & LaGanke, J. S. (2002). Computer-mediated communication and group decision making: A meta-analysis. *Organizational Behavior and Human Decision Process*, *87*, 156–179.

Barry, B., & Stewart, G. L. (1997). Composition, process, and performance in self-managed groups: The role of personality. *Journal of Applied Psychology*, *82*, 62–78.

Beaubien, J. M., & Baker, D. P. (2003). Post-training feedback: The relative effectiveness of team- versus instructor-led debriefs. In *Proceedings of the 47th annual meeting of the Human Factors and Ergonomics Society* (pp. 2033–2036). Denver, CO: Sage Publications.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–151). Washington, DC: American Psychological Association.

De Guinea, A. O., Webster, J., & Staples, D. S. (2012). A meta-analysis of the consequences of virtualness on team functioning. *Information & Management*, *49*, 301–308.

Driskell, J. E., Radtke, P. H., & Salas, E. (2003). Virtual teams: Effects of technological mediation on team performance. *Group Dynamics: Theory, Research, and Practice*, *7*, 297–323.

Dwyer, D. J., Oser, R. L., Salas, E., & Fowlkes, J. E. (2000). Performance measurement in distributed environments: Initial results and implications for training. *Military Psychology*, *11*, 189–215.

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., & Aloni-Zohar, M. (2009). The effect of accuracy of performance evaluation on learning from experience: The moderating role of after-event reviews. *Journal of Applied Social Psychology*, *39*, 541–563.

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

Flavell, J. E. (1979). Metacognition and cognitive monitoring: A new area of cognitive development inquiry. *American Psychologist*, *34*, 906–911.

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

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

Jarrett, S. M., Glaze, R. M., Arthur, W., Jr., Schurig, I., Villado, A. J., & Bennett, W., Jr. (2010, April). *The role of sex composition in team training performance*. Paper presented at the 25th annual meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Kozlowski, S. W. J., & Bell, B. S. (2003). Work groups and teams in organizations. In W. C. Borman, D. R. Ilgen & R. J. Klimoski (Eds.), *Handbook of psychology: Vol. 12. Industrial and organizational psychology* (pp. 333–375). London, UK: Wiley.

Kring, J. P. (2004). *Communication modality and after action review performance in a distributed immersive virtual environment* (Unpublished doctoral dissertation). University of Central Florida, Orlando.

Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement*, *23*, 127–135.

Meliza, L. L., Bessemer, D. W., & Hiller, J. H. (1994). Providing unit training feedback in the distributed interactive simulation environment. In R. F. Holz, H. Hiller, & H. H. McFann (Ed.), *Determinants of effective unit performance: Research on measuring and managing unit training readiness* (pp. 257–280). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Mesmer-Magnus, J. R., DeChurch, L. A., Jimenez-Rodriguez, M., Wildman, J., & Shuffler, M. (2011). A meta-analytic investigation of virtuality and information sharing in teams. *Organizational Behavior and Human Decision Processes*, *115*, 214–225.

Oden, K. B. (2009). *Distributed team training: Effective team feedback* (Unpublished doctoral dissertation). University of Central Florida, Orlando.

Prince, C., Salas, E., Brannick, M., & Orasnau, J. (2005). Beyond facilitation: An improved CRM debrief for safety training. *Human Factors and Aerospace Safety*, *5*, 1–22.

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology*, *8*, 197–208.

Salas, E., Oser, R. L., Cannon-Bowers, J. A., & Daskarolis-Kring, E. (2002). Team training in virtual environments: An event based approach. In M. K. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications. Human factors and ergonomics* (pp. 873–892). Mahwah, NJ: Erlbaum.

Sanchez-Ku, M. L., & Arthur, W., Jr. (2000). A dyadic protocol for training complex skills: A replication using female participants. *Human Factors*, *41*, 512–520.

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, 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.

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). Texas A&M University, College Station.

Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology*, *59*, 623–664.

Sundstrom, E., De Meuse, K. P., & Futrell, D. (1990). Work teams. *American Psychologist*, *45*, 120–133.

Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do team and individual debriefs enhance performance? A meta-analysis. *Human Factors*, *55*, 231–245.

Townsend, A. M., Demarie, S. M., & Hendrickson, A. R. (1998). Virtual teams: Technology and the workplace of the future. *Academy of Management Executive*, *12*, 17–29.

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

Villado, A. J., & Arthur, W., Jr. (2013). The comparative effectiveness of subjective and objective after-action reviews on team performance on a complex task. *Journal of Applied Psychology*, *98*, 514–528.

Williams, M., & Watson, A. (2004). Post-lesson debriefing: Delayed or immediate? An investigation of student teacher talk. *Journal of Education for Teaching*, *30*, 85–96.

Yorio, P. L., & Wachter, J. K. (2014). The impact of human performance focused safety and health management practices on injury and illness rates: Do size and industry matter? *Safety Science*, *62*, 157–167.