Do Team and Individual Debriefs Enhance Performance? A Meta-Analysis

Scott I. Tannenbaum and **Christopher P. Cerasoli**, the Group for Organizational Effectiveness, Albany, New York

Objective: Debriefs (or "after-action reviews") are increasingly used in training and work environments as a means of learning from experience. We sought to unify a fragmented literature and assess the efficacy of debriefs with a quantitative review.

Background: Used by the U.S. Army to improve performance for decades, and increasingly in medical, aviation, and other communities, debriefs systematize reflection, discussion, and goal setting to promote experiential learning. Unfortunately, research and theory on debriefing has been spread across diverse disciplines, so it has been difficult to definitively ascertain debriefing effectiveness and how to enhance its effectiveness.

Method: We conducted an extensive quantitative meta-analysis across a diverse body of published and unpublished research on team- and individual-level debriefs.

Results: Findings from 46 samples (N = 2,136) indicate that on average, debriefs improve effectiveness over a control group by approximately 25% (d = .67). Average effect sizes were similar for teams and individuals, across simulated and real settings, for within- or between-group control designs, and for medical and nonmedical samples. Meta-analytic methods revealed a bolstering effect of alignment and the potential impact of facilitation and structure.

Conclusion: Organizations can improve individual and team performance by approximately 20% to 25% by using properly conducted debriefs.

Application: Debriefs are a relatively inexpensive and quick intervention for enhancing performance. Our results lend support for continued and expanded use of debriefing in training and in situ. To gain maximum results, it is important to ensure alignment between participants, focus and intent, and level of measurement.

Keywords: experiential learning, teams, after-action review, team training, group dynamics, feedback, organizational learning, group performance

Address correspondence to Scott I. Tannenbaum, The Group for Organizational Effectiveness, 727 Waldens Pond Rd., Albany, NY 12203; e-mail: scott.tannenbaum@groupoe.com.

HUMAN FACTORS
Vol. 55, No. 1, February 2013, pp. 231-245
DOI:10.1177/0018720812448394

INTRODUCTION

For decades, the U.S. military has deployed after-action reviews (or "debriefs") designed to improve learning and performance (Morrison & Meliza, 1999). Debriefs lead individuals or teams through a series of questions that allow participants to reflect on a recent experience, construct their own meaning from their actions, and uncover lessons learned in a nonpunitive environment. Debriefs have become a common tool for supporting experiential learning in military settings and are becoming more common in other sectors as well. They are used in training settings (e.g., after a simulation exercise) as well as in situ (e.g., after a work experience).

Unfortunately, the theoretical and empirical literature examining debriefing is fragmented. The theoretical foundation is somewhat sparse, and the research examining debrief effectiveness is scattered across several disciplines, including medical, educational, psychological, and organizational fields. Researchers and practitioners lack a common point of reference from which to examine debriefs, establish their efficacy, and determine how they can best be deployed. In this article, we examine the efficacy of debriefs and what moderates their effectiveness. We first establish the defining elements of a debrief, providing an operational definition for inclusion in our subsequent meta-analysis. We then propose five hypotheses and two research questions regarding the efficacy of debriefs, followed by pertinent meta-analytic results. Finally, we conclude with a few recommendations for application and future research.

What Is a Debrief and Why Are They Effective?

Like other interventions designed to promote employee and team development (D'Abate, Eddy, & Tannenbaum, 2003), some conceptual ambiguity surrounds the current definition of a debrief. Foundationally, debriefing draws on principles from several areas of science (Ellis & Davidi, 2005), including information feedback, performance measurement, cognition and memory, group processes, communication theory, and instructional science. In practice, debriefing is discussed in the medical (Gaba, Howard, Fish, Smith, & Sowb, 2001), military (Ron, Lipshitz, & Popper, 2002), safety (Allen, Baran, & Scott, 2010), teamwork (Smith-Jentsch, Cannon-Bowers, Tannenbaum, & Salas, 2008), aerospace (Rogers & Milam, 2004), educational (Ellis, Ganzach, Castle, & Sekely, 2010), organizational training (Garvin, Edmondson, & Gino, 2008), human factors (Rosen, Salas, Tannenbaum, Pronovost, & King, 2012), and aviation (Dismukes & Smith, 2000) communities. Given the heterogeneous foundation and application of debriefs, some definitional ambiguity is not surprising.

Construct clarity is essential for any meta-analysis (cf. Le, Schmidt, Harter, & Lauver, 2010; Viswesvaran & Ones, 1995), so to conduct this meta-analysis, we must first clarify the essential elements that differentiate debriefs from other interventions. An intervention that lacks any of these elements, although potentially valuable, would not be considered a true debrief and would be excluded from the meta-analysis. After an extensive review of the literature, we identified four essential elements as shown in Table 1 and described next.

Active (vs. passive) self learning. Debriefs are fundamentally a form of emergent learning (Darling & Parry, 2001) in which individuals use an iterative process of reflection and planning to improve performance. Personal and active engagement produces a different type of insight than do more passive experiences (Ron et al., 2002). Self-discovery has been shown to be a key factor in effective developmental experiences (Eddy, D'Abate, Tannenbaum, Givens-Skelton, & Robinson, 2006), and active learning techniques also push the learners to engage in experimentation with ideas and actions, which reinforces cycles of learning (Kolb, 1984). As such, to be considered a true debrief, some *active* self-learning and self-discovery on the part of participants must take place (Ellis & Davidi, 2005). For example, simply receiving feedback about an experience or being told by a coach how to improve is *passive* and thus would not meet our operationalization of a debrief.

Developmental (vs. administrative) intent. Debriefs are intended primarily to serve developmental purposes rather than evaluative or judgmental purposes. A developmental, nonpunitive focus not only yields more accurate feedback but fosters an environment that encourages information exchange and perspective taking and maximizes experiential learning.

Research on performance appraisal sheds light on this dynamic. The purpose for which performance is rated can have a substantial impact on both rating accuracy and on the acceptance of feedback (Jawahar & Williams, 1997; Murphy & Cleveland, 1995). When raters know that administrative decisions will be made on the basis of their assessments, their ratings are typically less accurate. Moreover, when ratees believe the ratings can have punitive implications, they are more likely to be defensive, less willing to share information, and less open to feedback.

Given the importance of accurate and credible feedback and information sharing for enhancing motivation, learning, and performance (Kluger & DeNisi, 1996; Mesmer-Magnus & DeChurch, 2009), developmental intent should be considered a critical element of a debrief. An intervention that focuses primarily on evaluation, such as a performance review, would not be considered a debrief. Similarly, we would not consider incidental learning from experience (cf. Marsick & Watkins, 1990) to be a debrief.

Specific (vs. general) events. Part of what defines a debrief is a focus on specific activities, episodes, and events rather than on general performance or competencies. Reflecting on specific past events provides a different degree of focus and allows for a deeper examination of particular actions, cue-strategy associations, underlying cognitions, and so on than does a general discussion of overall performance. It also allows for the establishment of specific, challenging action plans and goals (Locke & Latham, 1990), which in turn affect motivational direction, intensity, and persistence (Locke & Latham, 2002). Therefore, general discussions surrounding team or individual strengths and weaknesses (e.g., receiving 360° feedback without focusing on a particular event, task, episode, or situation), would not be considered a debrief.

TABLE 1: Debriefs: Essential Elements

Element	Definition	Excludes
Active self-learning	Participants engage in some form of self-discovery or active involvement and are not merely passive recipients	Passive receipt of feedback; being told how to improve by a coach or facilitator
Developmental intent	A clear, primary intent for improvement or learning that is nonpunitive rather than judgmental or administrative	Performance appraisals or reviews; incidental learning
Specific events	Involves reflection on specific events or performance episodes rather than general performance or competencies	General discussion of a team's or person's overall strengths/weaknesses; 360° feedback about overall competencies
Multiple information sources	Includes input from multiple team members or from a focal participant and at least one external source, such as an observer or objective data source	Personal diary keeping; self-reflection

Multiple (vs. single) information sources. Multiple information sources are a critical element of a debrief for several reasons. Multiple sources increase the domain of coverage, yielding more diverse and complete accounts of a recent performance episode. Moreover, multiple sources of information can improve the credibility of feedback. According to goal-setting theory (Locke & Latham, 1990) and feedback intervention theory (Kluger & DeNisi, 1996), when feedback is perceived as credible, it yields better outcomes. For the purposes of our meta-analysis, a debrief cannot be a solitary experience. It must involve either a team sharing members' perceptions or a focal participant and at least one external information source, such as an observer, facilitator, and/or objective data source (e.g., video). So, for example, keeping a diary for self-reflection would not, in and of itself, be considered a debrief.

Summary. Debriefs, as defined in the preceding paragraphs, should be effective because they encourage reflection and self-discovery, target potential opportunities for improvement, and thus improve the quality and rate of experiential learning. By pairing active learning with multiple information sources to improve situational understanding and by identifying lessons learned and establishing specific future plans and goals, debriefs are designed to give individuals and teams a systematic, credible method for improving their performance. Thus we propose the following:

> *Hypothesis 1:* Individuals and teams that use debriefing are more effective than individuals and teams that do not use debriefing.

Moderators of Debrief Effectiveness

We hypothesize that several methodological and substantive factors should moderate the observed effectiveness of a debrief.

Alignment of levels. It is important to consider the level being examined in any debriefing study because the processes that result in team performance may not be the same as those involved in individual performance (Marks, Mathieu, & Zaccaro, 2001). Conceptually, there are three distinct levels to consider in any debrief. At the highest level (**participant level**), a debrief can be conducted either with a team or with an individual as the participant(s). Debriefs with teams are common in military and medical applications, whereas individual debriefs are more common in educational settings. The second level (*focal level*) is whether the debrief is focused primarily on improving the team as a whole or on independently improving each individual. The third level (*measurement level*) is more methodological in nature and involves considering whether the study measured performance at the individual level or at the team level.

As with other areas in psychology (cf. Ajzen & Fishbein, 1977), we suggest that the observed effectiveness of debriefs will be greater when the levels are aligned. For example, when the participants are a team and the focus is on enhancing team performance rather than the individual performance of team members, the discussion is more likely to directly explore critical teamwork processes, and the observed validity of the debrief should be higher. Similarly, when the focus is on improving team performance and performance metrics are collected at the team level, there is greater methodological alignment. In sum, alignment should reinforce the intent of the debrief and result in greater observed debrief effectiveness. Thus, we propose the following:

> *Hypothesis 2:* Debriefs are more effective when levels are aligned.

Optimal debrief characteristics. Three characteristics of a debrief can influence debrief effectiveness. First, *facilitation* is thought to improve the effectiveness of a debrief, because facilitation enhances objectivity and concentration. An impartial facilitator improves debrief effectiveness not only by streamlining task and interpersonal processes (cf. Dennis & Wixom, 2001) but by guiding individuals to consider the correct questions, avoiding entrapment in irrelevant details, drawing out participation by all team members, and ensuring that goals are set in an optimal manner.

Second, greater *structure* should improve the effectiveness of a debrief by enhancing focus, ensuring efficient learning processes, and increasing the quality of the discussion. Employing a highly structured approach that guides participants through the review and goal-setting process can free up the cognitive resources of participants, in essence allowing them to focus on reflecting, discussing, and learning from their experiences without worrying about managing the "process" (cf. Bargh & Chartrand, 1999). Increased structure can also help ensure the exchange of important information among team members (Stasser, Taylor, & Hanna, 1989).

Finally, it has become increasingly common to use some sort of *multimedia aid* during debriefing (e.g., video). It has often been suggested that aids can improve debrief quality (Ellis et al., 2010; Morrison & Meliza, 1999) because they permit participants and facilitators to more thoroughly dissect and understand a performance episode, painting a more accurate and actionable picture of what recently occurred. As such, we offer the following three hypotheses:

> *Hypothesis 3:* Facilitated debriefs are more effective than nonfacilitated debriefs.
>
> *Hypothesis 4:* A higher level of structure is associated with a more effective debrief.
>
> *Hypothesis 5:* Debriefs using multimedia aids are more effective than debriefs not using multimedia aids.

Methodological moderators. It is also important to examine the extent to which observed debriefing efficacy may be a function of methodological factors. We offer research questions for two potential methodological moderators: *study design* and *task environment*.

First, any experimental or quasiexperimental research that aims to make causal inferences requires some standard of control or comparison that enables isolation of the effect (cf. Shadish, Cook, & Campbell, 2002). Studies can be classified according to whether the comparison was assessed between groups or within groups. Although less desirable from a design perspective (cf. Cook, Campbell, & Peracchio, 1990), within-group controls are frequently the only available alternative because in applied contexts (when random assignment is often impossible), team sizes can be extremely uneven, groups may lack independence, and in some cases, only one group receives the intervention. Studies that employ between-group designs are generally considered more rigorous and thus may yield different results than those based on within-group designs.

> *Research Question 1:* Do within-group designs yield different observed validities than those that employ between-group designs?

A second factor that could affect the observed validity of debriefing is whether the task environment is simulated or real. Simulated tasks could have higher validities because a controlled setting has less error variance or "noise." For example, trigger events can be built into scenarios to create very specific experiences (Salas, Rhodenizer, & Bowers, 2000), promoting targeted learning and improvement. However, it would be equally plausible to reason that a lower level of fidelity in simulations could make them less realistic and could reduce participants' motivation to learn (cf. Goldstein & Ford, 2002). This point leads us to the following research question:

> *Research Question 2:* Does the task environment (i.e., simulated vs. real) affect the observed validity of a debrief?

METHOD

Literature Search

Multiple electronic databases were searched for literature containing variants of the words *debrief* or *after-action review* coupled with *performance, effectiveness, ratings*, and similar terms. Major databases searched included PsycINFO, Scopus, PubMed, ERIC, Military & Government Collection, Business Source, and MEDLINE, as well as an extensive collection of less well-known or less directly relevant databases from EBSCO (available from the second author by request). Our search returned 1,561 unique, nonduplicated references. We then carefully read the abstracts of all references to determine whether there was a chance of a reported relationship of interest, resulting in 218 published and 12 unpublished articles for review. At a minimum, articles had to include (a) measurement of performance before and after debriefing, (b) usable data with which to calculate an effect size, and (c) the presence of a debrief comprising the four defining elements discussed earlier.

Although debriefs have been used for multiple outcomes (e.g., safety climate, attitudes), we restricted our analyses to performance. Performance included objectively quantifiable output (e.g., simulator reports, game scores), personnel records (e.g., hospital records), self-ratings, and performance ratings (e.g., subject matter expert ratings, performance appraisal ratings). Studies reporting a clinical or critical incident stress debriefing were excluded, as these referred to exit counseling. Because our focus is primarily on organizational settings, we excluded any samples with respondents younger than 16. This resulted in a final calculation and inclusion of 111 effect sizes from 46 independent samples in 31 studies (29 published, 2 unpublished) with a total N of 2,136. We coded all studies' effect sizes and key characteristics using a structured coding guide. All articles were coded by one author, and a subset of articles was coded by both authors, yielding an agreement rate greater than 90%. Any conceptual disagreements or questions that arose during coding were discussed, and consensus was reached.

Hypothesized moderators were coded for each study. Levels of analysis were coded as defined in the Introduction. A study was considered to be facilitated if information suggested that a knowledgeable, impartial facilitator organized and conducted the debrief session. When information was available, level of structure was coded as high (protocol specified exact questions and procedures), moderate (protocol provided specific goals/objectives, allowing flexibility in deployment), low (protocol specified only general aims of the debrief), or none (protocol was explicitly nonexistent).

We also coded moderators to address the two research questions. Studies that relied solely on a within-unit (person, team) comparison were coded as within-group designs, whereas studies that compared the target group with a no-treatment control group were coded as between-group designs. The use of multimedia aids was considered present if some form of computer or electronic audio-video feedback was used during the debrief for the purposes of review. Other miscellaneous moderators that were coded for included population sampled from (medical vs. nonmedical), environmental setting (training simulation vs. real deployment), type of criteria (objective vs. subjective), sample percentage female, and time spent in debrief (averaged if there were multiple debriefs in one sample).

Analysis

We used the random-effects meta-analytic methods of Hunter and Schmidt (2004) to aggregate effect sizes from primary data up to the population level. Unlike fixed-effects models, random-effects models are built for generalizability to a larger population. The fundamental assumption in such models is that even the most comprehensive collection of studies for a meta-analysis can at best be considered a representative sample of the unobservable larger population one wishes to make inferences about. In contrast, fixed-effects models are built to describe the current sample. They assume that after correction for artifactual error, the same effect size underlies all studies (Hunter & Schmidt, 2004). Specifically, after correction for artifactual error, any additional variance must be attributable to moderators. The presence of additional variance is estimated through chi-square significance tests, which for a number of reasons can lead to inflated Type I (Hunter & Schmidt, 2000) or Type II (National Research Council, 1992, p. 52) error rates. Given the substantive issues and restrictive assumptions surrounding significance tests in meta-analysis and that this meta-analysis is designed to provide research conclusions that reach beyond the sample of studies observed (cf. Overton, 1998), we follow the call of previous research to use random effects models (Erez, Bloom, & Wells, 1996).

We used a conservative approach, in which the only correction for artifactual error applied was for sampling error. Reported statistics (e.g., t values, F tests, means and standard deviations) were converted to Cohen's d, a standardized estimate of the difference between debrief and control conditions in standard deviation units. To meet assumptions of independence, a sample could contribute only once to any meta-analytic estimate. In cases that reported multiple levels of a categorical moderator, superordinate overall analyses were collapsed across subordinate levels of the moderator so that they contributed only once. Corrections for criterion unreliability were not made for three main reasons. First, not many studies provided criterion reliability estimates, and when reported, the type of estimate was often unclear. Second, although statistically straightforward, applying a correction to interrater reliabilities of supervisor ratings is conceptually problematic because it assumes that any disagreement is attributable to error and not to different perspectives on performance (Murphy, 2008). Finally, because debriefs are an applied intervention, uncorrected effect sizes provide decision makers a more realistic idea of the degree of impact they can expect to observe when using debriefs (versus the "true" underlying theoretical relationship).

RESULTS

A review of the results revealed that a few studies demonstrated larger effect sizes, so we created a scatterplot of effect sizes against sample size to assess data patterns. Three studies were flagged as potential outliers (d = 4.71, 2.83, 3.81, respectively: O'Donnell et al., 2010; Qudrat-Ullah, 2007; and Smith-Jentsch et al., 2008). Removing those studies from the analysis reduced observed variance, skewness, and kurtosis. However, the three studies did not demonstrate any discernible pattern that could explain their larger effect sizes, either methodologically or substantively. Moreover, the overall pattern of results was the same for the full sample as for the trimmed sample. Therefore, in the tables, we report the results for both the full and trimmed samples but focus our attention primarily on the full sample of studies.

Table 2 presents the results related to Hypotheses 1 and 2, and Table 3 presents results related to the remaining hypotheses and research questions. In each table, we report the number of units (teams, individuals, or the smallest unit that preserved statistical independence), number of studies, and the effect size with its standard deviation corrected only for sampling error. We also include the percentage of variation in observed effect sizes attributable to sampling error, with a higher number indicating that

TABLE 2: Meta-Analysis of Debriefs and Performance

	Full Sample							Trimmed Sample (Outliers Removed)							
						File Drawer								80% CrI	
Variable	*n*	*k*	%↑	d_{obs}	SD_{obs}	.50	.20	*n*	*k*	%↑	d_{obs}	SD_{obs}	% err	.10	.90
Overall	2,136	46	25	0.67	.61	16	108	2,026	43	21	.54	.15	79.21	.35	.73
Participant level: Team															
Focal level: Team	546	16	25	0.66	.92	5	37	514	14	17	.45	.35	44.23	.00	.90
Measurement level: Team	176	10	38	1.20	1.37	14	50	144	8	21	.55	.21	79.64	.28	.83
Measurement level: Individual	370	6	16	0.41	.38	-1	6	370	6	16	.41	.38	31.45	–.08	.90
Participant level: Individual	1,255	23	26	0.71	.52	10	59	1,177	22	22	.57	.00	—	.57	.57
Focal level: Individual															
Measurement level: Individual	1,167	19	26	0.71	.55	8	48	1,089	18	21	.56	.00	—	.56	.56
Focal level: Team															
Measurement level: Individual	42	2	15	0.39	.00	0	2	42	2	15	.39	.00	—	—	—
Focal level: No information	46	2	33	0.94	—	2	7	46	2	33	.94	—	—	—	—

Note. All analyses are reported with and without three outliers omitted. *n* = number of individuals/teams/groups (analyses are presented at the smallest level of analysis possible to ensure sample independence); *k* = number of independent samples; %↑ = percentage increase in performance associated with using a debrief; d_{obs} = observed Cohen's *d* after removing sampling error; SD_{obs} = standard deviation after removing sampling error; file drawer = the number of unlocated studies that would have to arise reporting nonsignificant findings to drop our estimate below an effect size of .50 or .20; % err = percentage of variance in the observed Cohen's *d* accounted for by sampling error; 80% CrI = the lower and upper bounds of the 80% credibility interval for d_{obs}.

TABLE 3: Moderators of Debriefs and Performance

	Full Sample							Trimmed Sample (Outliers Removed)							
						File Drawer								80% CrI	
Moderator	n	k	%↑	d_{obs}	SD_{obs}	.50	.20	n	k	%↑	d_{obs}	SD_{obs}	% err	.10	.90
Facilitated															
Yes	1,350	34	27	.75	.76	17	94	1,240	31	20	.53	.17	78.45	.32	.75
No	53	2	10	.25	—	–1	1	53	2	10	.25	—	—	—	—
No information	760	10	21	.55	.14	1	18	760	10	21	.55	.14	74.18	.37	.73
Structured															
High	838	13	25	.69	.68	5	32	806	11	21	.56	.05	95.04	.49	.63
Medium	471	18	21	.54	.08	1	31	471	18	21	.54	.08	95.82	.44	.65
No	1	1	13	.32	—	0	1	1	1	13	.32	—	—	—	—
No information	826	14	27	.73	.70	6	37	748	13	20	.51	.23	56.47	.22	.81
Aids															
Yes	1,868	33	25	.66	.61	11	76	1,771	31	20	.53	.15	77.41	.34	.72
No	145	6	23	.62	.24	1	13	145	6	23	.62	.24	66.89	.31	.93
No information	123	7	34	.99	.86	7	28	110	6	25	.68	.17	89.78	.47	.89
Fidelity															
Simulated	1,788	40	26	.70	.67	16	100	1,678	37	20	.53	.15	80.28	.34	.73
Real	348	6	21	.56	.16	1	11	348	6	21	.56	.16	65.81	.35	.76
Control															
Within group	1,053	29	24	.63	.54	8	62	1,034	28	21	.56	.18	77.60	.33	.78
Between group	1,083	17	26	.72	.69	7	44	992	15	20	.52	.11	83.14	.37	.66

Note. All analyses are reported with and without three outliers omitted. n = number of individuals/teams/groups (analyses are presented at the smallest level of analysis possible to ensure sample independence); k = number of independent samples; %↑ = percentage increase in team performance associated with using a debrief; d_{obs} = observed Cohen's d after removing sampling error; SD_{obs} = standard deviation after removing sampling error; file drawer = the number of unlocated studies that would have to arise reporting nonsignificant findings to drop our estimate below an effect size of .50 or .20; % err = percentage of variance in the observed Cohen's d accounted for by sampling error; 80% CrI = the lower and upper bounds of the 80% credibility interval for d_{obs}.

variability in effect sizes can be attributed largely to natural variation of small samples. Finally, tables include the lower (.10) and upper (.90) bounds of the 80% credibility interval, which provides an estimate of the extent to which current findings are generalizable.

The mean publication year was 2009, indicating that the majority of the research was conducted within the past several years. On average, teams had an average of 5.25 participants with a slight male majority (47.84% female). An individual debrief session averaged 17.85 minutes to conduct. While single-session debriefs were not rare, the average number of debrief sessions reported was 3.28 across studies.

Overall, the use of debriefs resulted in an average d = .67, or a 25% improvement compared with control conditions, supporting Hypothesis 1, that debriefs improve performance (see Table 2). Findings revealed very similar effects whether the participants were teams (d = .66, or 25%) or individuals (d = .71, or 26%).

Hypothesis 2, that aligning levels would yield more effective debriefs, was supported. At the team level, when the focus of the debrief was team improvement, observed effectiveness was greater when team performance was measured (d = 1.20, or 38%) than when individual performance was measured (d = 0.41, or 16%). Because individual debriefs do not typically collect team performance data, we could not make an identical comparison at the individual level. However, for debriefs conducted with individual participants, improvements were greater when the focus of the debrief was to improve individual performance (d = 0.71, or 26%) rather than team improvement (d = 0.39, or 15%).

Hypotheses related to the debrief characteristics received mixed support (see Table 3). In line with Hypothesis 3, we found that facilitated debriefs (d = .75, or 27%) were about 3 times as effective as nonfacilitated debriefs (d = .25, or 10%) but the number of studies without facilitation was quite low. With regard to Hypothesis 4, that structure improves debrief effectiveness, the evidence was not conclusive. In the absence of any data for debriefs low in structure, we found that as debrief structure decreased from high to moderate to none, so too did its average effectiveness (d = .69, .54, .32, respectively), but the estimate for unstructured debriefs was based on only one study. Finally, Hypothesis 5 was not fully supported. Studies without multimedia aids (d = .66, or 25%) demonstrated similar results to those reporting no usage of multimedia aids (d = .62, or 23%). Our analyses also examined whether research design or task type affected debrief effectiveness (Research Questions 1 and 2). Compared with overall effectiveness (d = .67, or 25%), little variation was seen for within-group (d = .63, or 24%) or between-group controls (d = .72, or 26%). Modest differences were seen between debriefs conducted in simulated situations (d = .70, or 26%) and those in real settings (d = .56, or 21%). The small differences exhibited for both methodological factors disappeared in the trimmed sample.

Finally, we conducted several post hoc analyses exploring other potential moderators. We examined whether publication status moderated the observed validity of debriefs: If the average effect size for unpublished studies was lower, it might suggest the presence of a publication bias (i.e., that only studies reporting significant findings get published). Although we did not locate many, the unpublished studies we did find did not appear to be substantially different from the overall effect for published studies, and any differences were not in the direction to suggest publication bias (unpublished, k = 2, d = .77; published, k = 29, d = .67). In addition, we found that studies conducted in medical settings had similar results to those conducted in other settings (medical, k = 28, d = .66; nonmedical, k = 18, d = .69). Consistent with existing meta-analyses, studies that included subjectively rated criteria exhibited larger average effect sizes (d = 1.07) than did those that included objective criteria (d = 0.58). Approximately 54% of the studies employed objective criteria.

A few of our post hoc analyses also examined continuous moderators. Although weighted least squares (WLS) regression has been recommended for analyzing continuous moderators in meta-analysis (Steel & Kammeyer-Mueller, 2002), WLS was not possible because of missing data points. Instead, we used bivariate correlation to explore continuous moderators (as advocated by Hunter and Schmidt, 2004, p. 390, in these types of situations) and found that publication year ($k = 46$, $r = .02$, *ns*), sample size ($k = 46$, $r = -.20$, *ns*), percentage female ($k = 26$, $r = -.28$, *ns*), mean team size ($k = 11$, $r = -.08$, *ns*), and length of debrief session ($k = 17$, $r = .08$, *ns*) bore no statistically significant relationship to effect size.

DISCUSSION

Debriefs are a potentially powerful yet simple tool to improve the effectiveness of teams and individuals (Ron et al., 2002), but research and theory have been scattered across multiple disciplines. Moreover, although there have been more than 30 empirical studies that examined debriefing (containing 46 independent samples), no quantitative integration of the research existed with which to gauge debriefing efficacy. We attempted to overcome these shortcomings by clarifying essential debrief elements and establishing a set of research hypotheses and questions that could be examined through a meta-analytical review.

Our meta-analysis indicates that on average, debriefs improve performance by approximately 25%. Even excluding the three largest effect sizes yields a conservative average improvement of 21%. Pragmatically, an improvement of 20% or more is quite encouraging for an inexpensive intervention that requires little time to conduct (the average debrief studied lasted approximately 18 min). Moreover, debriefs appeared to work equally well for teams as they did for individuals.

Our findings indicate that aligning participants, intent, and measurement yield the greatest effects. When the goal is to improve *team* effectiveness, it makes sense to conduct debriefs with teams, to focus on improving the team, and to measure the performance of the team as a whole. In fact, on average, team debriefs that were conducted and studied in that manner showed an average effect size of $d = 1.20$, or 38%, including two of the three largest effect sizes in the meta-analysis. Similarly, when the goal is to improve *individual* effectiveness, focusing on improving the individual's performance (rather than the team's performance) is more effective. We should note however, that although alignment is clearly optimal, even "misaligned" debriefs demonstrated a reasonable level of efficacy, suggesting a broad range of acceptable applications.

We also examined three characteristics that are widely thought to improve the quality of debriefs: facilitation, structure, and multimedia aids. Unfortunately, we cannot definitively reach conclusions about these. It appears that facilitation helps, but the sample size for unfacilitated debriefs was too small to fully remove ambiguity. The pattern of results generally supported the hypothesis that structure enhances debrief effectiveness, but almost all studies involved debriefs with at least moderate structure, so there was insufficient variability to test that hypothesis conclusively.

Interestingly, the use of multimedia aids did not show a meaningful improvement in debriefing effectiveness. That finding does not mean that aids such as videotaping are never useful. Research studies generally receive extra attention, so it is possible that many of the debriefs studied were well designed and well facilitated and thus were less likely to benefit from or need multimedia aids to ensure their success. Similar to findings in research on web-based training (Sitzmann, Kraiger, Stewart, & Wisher 2006), the design of the intervention may be more important than the media employed, although in the future, researchers should examine the circumstances in which aids such as videotaping may be beneficial.

Overall, the efficacy of debriefs appears quite robust. Effect sizes averaged in the range of 20% to 26%, showing similar results across teams and individuals, across simulated and real settings, for within- or between-group control designs, and for medical and nonmedical participants. Moreover, there was no observable relationship between effect size and publication year, gender mix, time spent debriefing, or team size. Studies with objective criteria reported more conservative estimates of efficacy than did those with subjective criteria, but even those with objective criteria reported average improvements of approximately 20%.

Limitations

Any meta-analysis is only as robust as the studies it summarizes. Because much of the data were based on quasiexperimental designs, causal inferences should be drawn with care (Shadish et al., 2002). Although teams and individuals that debriefed demonstrated consistently better outcomes than did those in comparison conditions, a frequent lack of random assignment introduces ambiguity as to the isolation of causality. An alternative explanation is that teams selected for debriefing would have performed better regardless, but the lack of differences found for within-group and between-group controls tends to argue against that competing explanation. However, we cannot entirely rule out competing hypotheses because of the inclusion of studies with various degrees of design rigor.

Another limitation to the current review is a deficiency of certain data points. For example, the limited availability of studies with low structure or without a designated facilitator inhibits the strength of our conclusions regarding those characteristics. Moreover, the findings must be interpreted cautiously, because the meta-analysis did not include many unpublished studies. Unpublished studies are often more likely to report small or nonsignificant results, and as such, their exclusion (intentional or otherwise) may lead to inflated meta-analytic estimates. Nevertheless, in this meta-analysis, three indicators suggest that publication bias is unlikely to be a problem. First, a scatterplot of effect size against sample size showed no clear truncation of the distribution and approximated a desirable bell shape (versus an undesirable nonuniform distribution), which is inconsistent with the presence of publication bias (cf. Aguinis, Pierce, Bosco, Dalton, & Dalton, 2011). Second, our file-drawer analysis suggests that the findings regarding overall debrief efficacy are quite robust, requiring many statistically nonsignificant studies to nullify our findings. Finally, although inconclusive taken alone, a lack of difference between published and unpublished studies is also inconsistent with the presence of publication bias. Together, this pattern of evidence suggests that publication bias is unlikely in the current study, but we acknowledge that this threat cannot be ruled out definitively.

Applications for Practice

Debriefs have become increasingly common, as they are a relatively low-cost intervention. However, until now, the efficacy of debriefs had not been validated in a quantitative, integrated manner. Our findings show that debriefs yield 20% to 25% average improvements, which can be boosted by properly aligning how debriefs are conducted and perhaps through structure and effective facilitation. The results provide empirical support for continued and expanded use of both team and individual debriefs. Empirical evidence of this type may help encourage busy team leaders and trainers to allocate time for conducting structured debriefs.

Currently, team debriefs have been used most frequently in simulation training in military and, more recently, medical settings. Given their efficacy, we would suggest that debriefing should be a standard part of any team training intervention that incorporates a simulated team experience. In such cases, to promote alignment, the debrief should involve the full team, focus on team improvements, and assess effectiveness with team-level performance measures.

Organizations are looking for ways to promote on-the-job learning (Tannenbaum, Beard, McNall, & Salas, 2010). There is a clear opportunity to use debriefing as a way to promote team and individual learning from “real,” in situ experiences more frequently. In particular, situations in which a specific performance episode can be examined, such as a team assignment, sales call, safety effort, project launch, patient case, service experience, product development, or even a team meeting, would appear to be prime candidates for real-world debriefing.

Although we cannot say so unequivocally, debriefs appear to benefit from the provision of sufficient structure and guidance. Multimedia aids (e.g., video) may be one way of building in structure and guidance, but our analyses suggest that debriefs can be successful even without multimedia aids. When designing a debrief, we would advocate establishing a clear, structured process that incorporates the four foundation elements. Then, decide whether tools or aids are likely to help, and take actions to ensure that the person guiding the debrief is well prepared, as prior research has shown that not everyone is naturally ready to be an effective debrief facilitator (Dismukes, Jobe, & McDonnell, 2000).

Applications for Theory and Future Research

One theoretical contribution of the current study is the formal specification of the four defining elements. We treated these as criteria for inclusion in the meta-analysis, but in addition, they provide a theoretical foundation for differentiating debriefs from other interventions. Future researchers should corroborate or suggest modifications so that there is an accepted operational definition of debriefs with which to guide subsequent efforts.

Future research is needed to further explicate the factors that moderate the efficacy of debriefs. In particular, research should examine how various debriefing processes, techniques, and tools influence debrief efficacy. For example, when is greater structure needed, the use of videotaping advantageous, or the application of particular questioning techniques most beneficial? Additional moderators that were not adequately covered in prior research but of theoretical importance include temporal proximity of the review event, presence of advance-action review, and ad hoc versus intact teams. In addition, it could be beneficial to extend the criterion space beyond immediate learning and performance, for example, by examining whether debriefing with one team builds capabilities that individuals can use in subsequent team assignments.

The mediators of debriefing efficacy also need explication. In other words, why does debriefing boost performance? In team settings, does it operate by enhancing team processes (LePine, Piccolo, Jackson, Mathieu, & Saul, 2008)? Does it boost the formation of shared mental models (DeChurch & Mesmer-Magnus, 2010b) or collective efficacy (DeChurch & Mesmer-Magnus, 2010a)? Do individual debriefs help people acquire greater self-awareness, develop greater self-efficacy or motivation, or establish clearer goals? We now have a solid evidentiary basis for conducting debriefs; in the future, researchers can further refine our understanding of why they work and how best to deploy them.

CONCLUSION

For any developmental intervention to work, lessons must be integrated into everyday practice (D'Abate et al., 2003). The current meta-analysis shows that debriefs are effective (averaging 20–25% improvements), are efficient (averaging approximately 18 minutes, require few resources to conduct), and thus are worthy of organizational resources and support. Although the process has been used successfully in medical, educational, and military environments for 30 years, the current review suggests that usage could be expanded with substantial utility in a wide range of settings.

ACKNOWLEDGMENTS

This research was supported in part by Grant No. NNX11AR22G from the National Aeronautics and Space Administration.

KEY POINTS

- Debriefs are a quick, effective tool for improving team and individual performance.
- Meta-analytic results from 46 independent samples show that debriefs improve performance an average of 20% to 25%. Debriefs work equally well for teams and individuals.
- Debriefs work best when properly aligned: If the goal is to improve team performance, debriefs should be conducted with, measure, and focus on teams rather than individuals (and vice versa).
- Findings suggest that debriefs are even more effective when structured and facilitated.

REFERENCES

An asterisk indicates inclusion in the quantitative meta-analysis.

Aguinis, H., Pierce, C. A., Bosco, F. A., Dalton, D. R., & Dalton, C. M. (2011). Debunking myths and urban legends about meta-analysis. *Organizational Research Methods, 14*, 306–331.

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84*, 888–918.

Allen, J. A., Baran, B. E., & Scott, C. W. (2010). After-action reviews: A venue for the promotion of safety climate. *Accident Analysis & Prevention, 42*, 750–757.

*Anderson, J. M., Murphy, A. A., Boyle, K. B., Yaeger, K. A., & Halamek, L. P. (2006). Simulating extracorporeal membrane oxygenation emergencies to improve human performance: Part II. Assessment of technical and behavioral skills. *Simulation in Healthcare, 1*, 228–232.

*Andreatta, P., Saxton, E., Thompson, M., & Annich, G. (2011). Simulation-based mock codes significantly correlate with improved pediatric patient cardiopulmonary arrest survival rates. *Pediatric Critical Care Medicine, 12*, 33–38.

*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.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*, 462–479.

*Bethune, R., Govindarajulu, S., Sahu, A., Cawthorn, S., & Pullyblank, A. (2011). Use of briefings and debriefings as a tool in improving team work, efficiency, and communication in the operating theatre. *Postgraduate Medicine Journal, 87*, 331–334. doi:10.1136/pgmj.2009.095802

*Boet, S. (2011). Looking in the mirror: Self-debriefing versus instructor debriefing for simulated crises. *Critical Care Medicine, 39*, 1–5.

*Catchpole, K. R., Dale, T. J., Hirst, D. G., Smith, J. P., & Giddings, T. A. E. B. (2010). A multicenter trial of aviation-style training for surgical teams. *Journal of Patient Satisfaction, 6*, 180–186.

*Chung, S. P., Cho, J., Park, Y. S., Kang, H. G., Kim, C. W., Song, K. J., Lim, H., & Cho, G. C. (2010). Effects of script-based role play in cardiopulmonary resuscitation team training. *Emergency Medicine Journal, 28*, 690–694. doi:10.1136/emj.2009.090605

Cook, T. D., Campbell, D. T., & Peracchio, L. (1990). Quasi experimentation. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1., 2nd ed., pp. 491–576). Palo Alto, CA: Consulting Psychologists Press.

D'Abate, C. P., Eddy, E. R., & Tannenbaum, S. I. (2003). What's in a name? A literature-based approach to understanding mentoring, coaching, and other constructs that describe developmental interactions. *Human Resource Development Review, 2*, 360–384.

Darling, M. J., & Parry, C. S. (2001). *From post-mortem to living practice: An in-depth study of the evolution of the after-action review*. Boston, MA: Signet Consulting Group.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010a). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology, 95*, 32–53.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010b). Measuring shared team mental models: A meta-analysis. *Group Dynamics: Theory, Research, & Practice, 14*, 1–14.

Dennis, A. R., & Wixom, B. H. (2001). Investigating the moderators of the group support systems use with meta-analysis. *Journal of Management Information Systems, 18*, 235–257.

*DeVita, M. A., Schaefer, J., Lutz, J., Wang, H., & Dongilli, T. (2005). Improving medical emergency team (MET) performance using a novel curriculum and a computerized human patient simulator. *Quality and Safety in Healthcare, 14*, 326–331.

*Dine, C. J., Gersh, R. E., Leary, M., Riegel, B. J., Bellini, L. M., & Abella, B. S. (2008). Improving cardiopulmonary resuscitation quality and resuscitation training by combining audiovisual feedback and debriefing. *Critical Care Medicine, 36*, 2817–2822.

Dismukes, R. K., Jobe, K. K, & McDonnell, L. K. (2000). Facilitating LOFT debriefings: A critical analysis. In R. K. Dismukes & G. M. Smith (Eds.), *Facilitation in aviation training and operations* (pp. 13–25). Aldershot, UK: Ashgate.

Dismukes, R. K., & Smith, G. M. (Eds.). (2000). *Facilitation in aviation training and operations*. Aldershot, UK: Ashgate.

*Donoghue, A., Ventre, K., Boulet, J., Brett-Fleegler, M., Nishisaki, A., Overly, F., & Cheng, A. (2011). Design, implementation, and psychometric analysis of a scoring instrument for simulated pediatric resuscitation: A report from the EXPRESS pediatric investigations. *Simulation in Healthcare, 6*, 71–77.

Eddy, E. R., D'Abate, C. P., Tannenbaum, S. I., Givens-Skeaton, S., & Robinson, G. (2006). Key characteristics of effective and ineffective developmental interactions. *Human Resource Development Quarterly, 17*, 59–84.

*Edelson, D. P., Litzinger, B., Aroroa, V., Walsh, D., Kim, S., Lauderdale, D. S., Vanden Hoeck, T. L., Becker, L. B., & Abella, B. S. (2008). Improving in-hospital cardiac arrest process and outcomes with performance debriefing. *Archive of Internal Medicine, 168*, 1063–1069.

*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 experience: The moderating role of kind of after-event review. *Journal of Applied Psychology, 91*, 669–680.

*Ellis, S., Mendel, R., & Zohar, D. (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.

Erez, A., Bloom, M. C., & Wells, M. T. (1996). Using random rather than fixed effects models in meta-analysis: Implications for situational specificity and validity generalization. *Personnel Psychology, 49*, 275–306.

Gaba, D. M., Howard, S. K., Fish, K. J., Smith, B. E., & Sowb, Y. A. (2001). Simulation-based training in anesthesia crisis resource management (ACRM): A decade of experience. *Simulation & Gaming, 32*, 175–193.

Garvin, D. A., Edmondson, A. C., & Gino, F. (2008, March). Is yours a learning organization? *Harvard Business Review, 89*, 1–11.

*Geis, G. L., Pio, B., Pendergrass, T. L., Moyer, M. R., & Patterson, M. D. (2011). Simulation to assess the safety of new healthcare teams and new facilities. *Simulation in Healthcare, 6*, 125–133. doi:10.1097/SIH.0b013e31820dff30

*Goffmann, D., Heo, H., Pardanani, S., Merkatz, I. R., & Bernstein, P. S. (2008). Improving shoulder dystocia management among resident and attending physicians using simulations. *American Journal of Obstetrical Gynecology, 199*, 294.e1–e5.

Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations* (4th ed.). Belmont, CA: Wadsworth Thomson Learning.

*Hamad, G. G., Brown, M. T., & Clavijo-Alvarez, J. A. (2007). Postoperative video debriefing reduces technical errors in laparoscopic surgery. *American Journal of Surgery, 194*, 110–114.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative knowledge in psychology. *International Journal of Selection and Assessment, 8*, 275–292.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis* (2nd ed.). Thousand Oaks: Sage.

Jawahar, I. M., & Williams, C. R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology, 50*, 905–925.

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

Kolb, D. (1984). *Experiential learning: Experience as a source of learning*. Englewood Cliffs, NJ: Prentice Hall.

*Kring, J. P. (2005). Communication modality and after action review performance in a distributed immersive virtual environment. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 66*, 598.

Le, H., Schmidt, F. L., Harter, J. K., & Lauver, K. J. (2010). The problem of empirical redundancy of constructs in organizational research: An empirical investigation. *Organizational Behavior and Human Decision Processes, 112*, 112–125.

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology, 61*, 273–307.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*, 705–717.

*Marken, P. A., Zimmerman, C., Kennedy, C., Schremmer, R., & Smith, K. V. (2010). Human simulators and standardized patients to teach difficult conversations to interprofessional health care teams. *American Journal of Pharmaceutical Educational, 74*, 1–8.

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

Marsick, V. J., & Watkins, K. E. (1990). *Informal and incidental learning in the workplace.* London, UK: Routledge.

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

*Mikrogianakis, A., Osmond, M. H., Nuth, J. E., Shepard, A., Gaboury, I., & Jabbour, M. (2008). Evaluation of a multidisciplinary pediatric mock trauma code educational initiative: A pilot study. *Journal of Trauma: Injury, Infection, and Critical Care, 64*, 761–767.

Morrison, J. E., & Meliza, L. L. (1999). *Foundations of the after action review process* (Special Report 42). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Murphy, K. R. (2008). Explaining the weak relationship between job performance and ratings of job performance. *Industrial and Organizational Psychology, 1*, 148–160.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives.* Thousand Oaks, CA: Sage.

National Research Council. (1992). *Combining information: Statistical issues and opportunities for research.* Washington, DC: National Academy of Sciences Press

*O'Donnell, J. M., Goode, J. S., Henker, R. A., Kelsey, S., Bircher, N., Peele, P., Bradle, J., Engberg, R., Close, J., & Sutton-Tyrrell, K. (2010). An ergonomic protocol for patient transfer that can be successfully taught using simulation methods. *Clinical Simulation in Nursing, 8*, e3–e14. doi:10.1016/j.ecns.2010.05.003

*Oden, K. B. (2009). Distributed team training: Effective team feedback. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 70*(1-B), 721.

Overton, R. C. (1998). A comparison of fixed effects and mixed (random effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods, 3*, 354–379.

*Qudrat-Ullah, H. (2007). Debriefing can reduce misperceptions of feedback: The case of renewable resource management. *Simulation Gaming, 38*, 382–397.

*Robertson, B., Schumacher, L., Gosman, G., Kanfer, R., Kelley, M., & DeVita, M. (2009). Simulation-based crisis team training for multidisciplinary obstetric providers. *Simulation in Healthcare, 4*, 77–83.

Rogers, E. W., & Milam, J. (2004). Pausing for learning: Applying the after action review process at the NASA Goddard Space Flight Center. In *Proceedings of the IEEEAC* (pp. 4383–4388). Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Ron, N., Lipshitz, R., & Popper, M. (2002). How organizations learn: Post-flight reviews in a F-16 fighter squadron. *Organization Studies, 27*, 1069–1089.

Rosen, M. A., Salas, E., Tannenbaum, S. I., Pronovost, P., & King, H. B. (2012). Simulation-based training for teams in healthcare: Designing scenarios, measuring performance, and providing feedback. In P. Carayon (Ed.), *Handbook of human factors and ergonomics in health care and patient safety* (pp. 573–594). Boca Raton, FL: CRC Press.

Salas, E., Rhodenizer, L. & Bowers, C. A. (2000). The design and delivery of crew resource management training: Exploiting available resources. *Human Factors, 42*, 490–511.

*Schwid, H. A., Rooke, G. A., Michalowski, P., & Ross, B. K. (2001). Screen-based anesthesia simulation with debriefing improves performance in a mannequin-based anesthesia simulator. *Teaching and Learning in Medicine, 13*, 92–96.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton, Mifflin.

Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology, 59*, 623–644.

*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.

Stasser, G., Taylor, L. A., & Hanna, C. (1989). Information sampling in structured and unstructured discussions of three- and six-person groups. *Journal of Personality and Social Psychology, 57*, 67–78.

Steel, P. D., & Kammeyer-Mueller, J. D. (2002). Comparing meta-analytic moderator estimation techniques under realistic conditions. *Journal of Applied Psychology, 87*, 96–111.

*Sutton, R. M., Niles, D., Meaney, P., Aplenc, R., French, B., Abella, B. S., Lengetti, E., Berg, R. A., Helfaer, M. A., & Nadkami, V. (2009). Abstract P55: Brief bedside CPR skill debriefing and automated corrective feedback during simulated pediatric cardiac arrest. *Circulation, 120*, s1453.

Tannenbaum, S. I., Beard, R. L., McNall, L. A., & Salas, E. (2010). Informal learning and development in organizations. In S. W. J. Kozlowski & E. Salas (Eds.), *Learning, training, and development in organizations* (pp. 303–332). New York, NY: Routledge.

*Villado, A. (2009). The after-action review training approach: An integrative framework and empirical investigation. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 69*, 6463.

Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology, 48*, 865–885.

*Welke, T. M., LeBlanc, V. R., Savoldelli, G. L., Joo, H. S., Chandra, D. B., Crabtree, N. A., & Naik, V. N. (2009). Personalized oral debriefing versus standardized multimedia instruction after patient crisis simulation. *Anesthesia & Analgesia, 109*, 183–189.

*Zausig, Y. A., Grube, C., Boeker-Blum, T., Busch, C. J., Bayer, Y., Sinner, B., Zink, W., Schaper, N., & Graf, B. M. (2009). Inefficacy of simulator-based training on anesthesiologists' non-technical skills. *Acta Anesthesiological Scandinavica, 53*, 611–619.

Scott I. Tannenbaum is the president of the Group for Organizational Effectiveness (gOE), an Albany, New York–based research and consulting firm. He

received his PhD in industrial-organizational psychology in 1986 from Old Dominion University.

Christopher P. Cerasoli is a doctoral candidate and research associate with the Group for Organizational Effectiveness (gOE). He received his master's in psychology from the State University of New York at Albany and is currently pursuing a PhD in industrial-organizational psychology.

Date received: November 8, 2011
Date accepted: April 20, 2012