Does Team Building Work?

Cameron Klein, Deborah DiazGranados, Eduardo Salas, Huy Le, C. Shawn Burke, Rebecca Lyons and Gerald F. Goodwin

Small Group Research 2009 40: 181 originally published online 6 January 2009
DOI: 10.1177/1046496408328821

Small Group Research
Volume 40 Number 2
April 2009 181-222

10.1177/1046496408328821
http://sgr.sagepub.com
hosted at
http://online.sagepub.com

Does Team Building Work?

Cameron Klein
Deborah DiazGranados
Eduardo Salas
Huy Le
C. Shawn Burke
Rebecca Lyons
University of Central Florida
Gerald F. Goodwin
Army Research Institute

This research reports the results of a comprehensive investigation into the effectiveness of team building. The article serves to update and extend Salas, Rozell, Mullen, and Driskell's (1999) team-building meta-analysis by assessing a larger database and examining a broader set of outcomes. Our study considers the impact of four specific team-building components (goal setting, interpersonal relations, problem solving, and role clarification) on cognitive, affective, process, and performance outcomes. Results (based on 60 correlations) suggest that team building has a positive moderate effect across all team outcomes. In terms of specific outcomes, team building was most strongly related to affective and process outcomes. Results are also presented on the differential effectiveness of team building based upon the team size.

Keywords: *team building; team performance; team development*

> Teams of people working together for a common cause touch all our lives. From everyday activities like air travel, fire fighting, and running the United Way drive to amazing feats of human accomplishment like climbing Mt. Everest and reaching for the stars, teams are at the center of how work gets done in modern life.
>
> Kozlowski & Ilgen, 2006, p. 78

This quote, from a recent review of work-team effectiveness, exemplifies the central role that teams play in our lives. Although many labels have been applied to team-based forms of organizing (i.e., crews, teams, groups, and collectives), these entities are essential to the accomplishment of organizational goals. Indeed, there is ample support in the literature for the

contention that team-based forms of organizing are beneficial both to organizations and to individuals. For example, Applebaum and Batt (1994) reviewed 12 large-scale surveys and 185 case studies of managerial practices and concluded that team-based work leads to improvements in organizational performance on measures of both efficiency and quality. They argued that team-based systems benefit workers because of the higher likelihood of job enhancement, autonomy, and skill development associated with these systems. However, the simple existence of a team-based organizing structure is not enough to ensure that positive outcomes will result. Teams must be nurtured, supported, and developed.

The Motivation for Understanding the Efficacy of Team Building

There are three motivations for understanding the efficacy of team-building interventions in organizations. First, team building is one of the most commonly applied group development interventions in organizations today. It is widely used and comes in many forms, including outdoor experiential activities and indoor group process discussions. However, no one is quite sure how and why these interventions work, or if they even work at all. Considering the vast sum of money directed toward the development of teams in organizations, it is important that practitioners (and researchers) gain a better understanding of the effectiveness and boundary conditions of team building.

Second, as there are many options available to organizations in the pursuit of improved teamwork, it is important to determine whether team building is a worthy choice. Some of these interventions are organizational

Authors' Note: This work was partially supported by funding from the U.S. Army Research Institute for the Behavioral and Social Sciences (Contract W74V8H-04-C-0025) and grant SES0527675 from the National Science Foundation, awarded to Glenn Harrison, Stephen M. Fiore, Charlie Hughes, and Eduardo Salas. All opinions expressed in this article are those of the authors and do not necessarily reflect the official opinion or position of the University of Central Florida, the Department of the Army, the Department of Defense, or the National Science Foundation. Portions of the article were presented at the 2nd Annual Conference of the Interdisciplinary Network for Group Research (INGRoup) held at Michigan State University, in Lansing, Michigan. Correspondence concerning this article should be addressed to: Dr. Eduardo Salas, Institute for Simulation & Training, University of Central Florida, 3100 Technology Parkway, Suite 132, Orlando, FL 32826, USA. Tel: (407) 882-1325; fax: (407) 882-1550; e-mail: esalas@ist.ucf.edu.

or structural in nature and do not specifically target team member interactions (e.g., job redesign, selection systems, and group incentive and performance management programs). In contrast, team-development interventions—those whose quest is to directly impact the functioning and effectiveness of work teams—provide the focus for the current research integration. These interventions, when properly conducted, can have a positive impact on organizations. Consider a research study conducted by Macy and Izumi (1993), who analyzed 131 studies of organizational change. They found that interventions with the largest effects upon financial measures of organizational performance were team-development interventions. That is, of all organizational interventions, those that focus on team development had the largest effect on measures of financial performance.

Third, beyond financial performance, it is widely understood that team developmental interventions are key mechanisms that may be used to facilitate team effectiveness (Noe, 2002). Therefore, it is important to understand how these interventions are most effective. At a general level, team-development interventions may include some form of team-training or team-building activities. Although both types of team-development interventions are designed to improve team functioning and effectiveness (especially when the science of individual and team training is utilized; Salas & Cannon-Bowers, 1997), team training and team building differ in important ways (Tannenbaum, Beard, & Salas, 1992). Team training is skill-focused (i.e., it is focused on gaining specific competencies), typically includes a practice component, and is done in context. It is generally formal and systematic. Team building, on the other hand, does not target skill-based competencies, is not systematic in nature, and is typically done in settings that do not approximate the actual performance environment. For our purposes, we define team building as a class of formal and informal team-level interventions that focus on improving social relations and clarifying roles, as well as solving task and interpersonal problems that affect team functioning. Team building works by assisting individuals and groups to examine, diagnose, and act upon their behavior and interpersonal relationships (Schein, 1969, 1999).

In light of the above, this article investigates the efficacy of team building. We begin with a brief review of conceptual issues and methodological issues in team building. Next, the hypotheses for this research are presented. Finally, the results of several meta-analytic integrations are discussed. First, however, is the rationale for the present research.

Understanding the Need for an Update

It is an unfortunate indictment of the literature and practice in this area that we are still searching for answers to questions posed by Beer (1976) and Salas, Rozell, Mullen, and Driskell (1999). Namely, does team building result in positive outcomes? Why? Under what conditions? Upon a careful review of the extent literature, it is clear that these questions need to be examined more closely to clarify our understanding of the effectiveness and boundary conditions of team building. Thus, this article reports the findings from a series of meta-analytic investigations of research on the efficacy of team-building interventions to update and extend the current state of knowledge in the team-building domain. In total, this research will examine three moderator variables in addition to making an overall assessment of the efficacy of team-building interventions. The current replication is meant to be systematic, rather than direct in nature (e.g., Aronson, Ellsworth, Carlsmith, & Gonzales, 1990). That is, rather than replicating the exact analyses from the study by Salas and colleagues, the current research will replicate and extend the earlier study in an effort to resolve ambiguities and provide new insights for organizational stakeholders and academicians alike concerning the effectiveness of team building.

Although the empirical evidence on team-building interventions is limited, a critical investigation of the available literature is warranted for several reasons. First, since the 1990s there has been an increasing incursion of team-building interventions in organizations. Some of the most recent trends in team building have taken these interventions into the kitchen and even the wilderness. Second, many practitioners feel that these interventions are useful. However, a more careful exploration of the utilities and strengths of these interventions would benefit practitioners now and would benefit the development of these interventions in the long run. Finally, there are enough data available to form estimates regarding some of the variables that may moderate the impact a team-building intervention may have on team outcomes.

There is a general consensus for the idea that the science of team training can be applied to enhance team functioning across organizations (e.g., Kozlowski & Ilgen, 2006; Salas & Cannon-Bowers, 1997). At the same time, reviews of team effectiveness have noted the inconsistent findings for the effectiveness of team-building interventions. Despite these inconsistent findings, researchers such as Kozlowski and Ilgen have acknowledged the potential that these interventions may have on shaping team development and on improving team effectiveness. Given the increasing frequency of

team-building interventions, if researchers do not catch up and start doing this research, an important window might be missed to constructively shape the important practice in this area. Thus, examining team building further—a topic that has not received recent sufficient attention in the literature—is critical. In the next section, a number of conceptual and methodological issues in team building are discussed.

Team Building

Conceptual Issues in Team Building

Originally designed as a group process intervention (e.g., Schein, 1969, 1999) for improving interpersonal relations and social interactions, team building has evolved to also include a concern for achieving results, meeting goals, and accomplishing tasks (Payne, 2001). In the late 1990s, Salas and colleagues (1999) described team-building interventions as extremely popular and common. According to Beer (1976), there are four basic approaches to team building, including: (a) a goal-setting, problem-solving model; (b) an interpersonal model; (c) a role model; and (d) the Managerial Grid (Blake & Mouton, 1964) model. However, this initial conceptualization has since been reconsidered.

Refinements to this four-pronged system began with the Managerial Grid model being dropped as a distinct team-building approach. In addition to dropping the Managerial Grid model, modern conceptualizations have separated the goal-setting and problem-solving approaches, using Buller's (1986) problem-solving component as a distinct approach. As discussed by Buller, team-building models rarely exist in pure form. That is, the interventions reported in the literature usually involve elements from several or all of the models. As a result, he proposed a general problem-solving model that follows Dyer's (1977) problem-solving framework. This model incorporates a focus on task or interpersonal issues, goal setting, and role clarification, depending on the nature of the specific problems identified for the group under investigation. Moreover, the problem-solving approach to team building is said to subsume each of Beer's (1976) components, and, as perhaps evident by its title, emphasizes the identification of major problems in the team. All told, these modifications have added clarity to the investigation and implementation of team building. Unfortunately, this newfound conceptual clarity may not have come soon enough for many investigators who had previously sought to assess the efficacy of team building.

Perhaps because of the conceptual confusion that existed in this area, many reviews of team building did not include the same articles. For example two noteworthy reviews (DeMeuse & Liebowitz, 1981; Woodman & Sherwood, 1980) had only 14 studies in common out of a total of 66 studies reviewed. A subsequent review by Buller (1986) included only 9 studies, 3 in common with Woodman and Sherwood (1980) and 6 in common with DeMeuse and Liebowitz. One reason for the apparent lack of consistency in the articles chosen for reviews is that team building has been an ill-defined concept (Buller, 1986). At the time of these early reviews there was no agreed-upon operational definition of the intervention. Taken together, there is now a consensus position that there exist four distinct models of team building. Although combinations of these approaches are common, the models include goal-setting, developing interpersonal relations, clarifying roles, and creating additional capacity for problem solving (Beer, 1976; Buller, 1986; Dyer, 1987; Salas et al., 1999). Table 1 describes each of the four models or components of team building in more detail.

Despite the recently established consensus on team-building components, there have been other problematic issues that have persisted for researchers and practitioners of this topic. Specifically, many early efforts were plagued by a number of methodological issues. A few of these are discussed in the next section.

Methodological Issues in Team Building

Early reviews of team building described both a lack of extensive research on the issue and trepidation concerning the methodological rigor of published studies (Buller, 1986; DeMeuse & Liebowitz, 1981; Tannenbaum et al., 1992; Woodman & Sherwood, 1980). That is, even if one could get beyond the disagreement concerning operational definitions of team building, much of the previous team-building research is characterized by methodological flaws (i.e., study design and measurement issues). For example, Buller (1982) found that more than half of the reported team-building studies employed pre-experimental designs—designs that do not allow for causal inferences. Also problematic, there was often no attempt to disentangle the effects of team building from other interventions that may be in process within an organization. Although this can be a common problem for field research in general, it has been particularly problematic in the team-building domain.

Tannenbaum and colleagues (1992) highlighted another methodological flaw in the team-building research. Specifically, they pointed out that there

Table 1
Models/Components of Team Building

Component	Salas, Rozell, Mullen, & Driskell, 1999	Salas, Priest, & DeRouin, 2005
Goal setting	Emphasis: Setting objectives and development of individual and team goals. Team members: Become involved in action planning to identify ways to achieve goals.	Designed to strengthen team member motivation to achieve team goals and objectives. By identifying specific outcome levels, teams can determine what future resources are needed. Individual characteristics (e.g., team member motivation) can also be altered by use of this intervention.
Interpersonal relations	Emphasis: Increasing teamwork skills (i.e., mutual supportiveness, communication, and sharing of feelings). Team members: Develop trust in one another and confidence in the team.	Based on the assumption that teams with fewer interpersonal conflicts function more effectively than teams with greater numbers of interpersonal conflicts. Requires the use of a facilitator to develop mutual trust and open communication between team members. As team members achieve higher levels of trust, cooperation, and cohesiveness, team characteristics can be changed as well.
Role clarification	Emphasis: Increasing communication among team members regarding their respective roles within the team. Team members: Improve understanding of their own and others' respective roles and duties within the team.	Defines the team as comprising a set of overlapping roles. These overlapping roles are characterized as the behaviors that are expected of each individual team member. Can be used to improve team and individual characteristics (i.e., by reducing role ambiguity) and work structure by negotiating, defining, and adjusting team member roles.
Problem solving	Emphasis: Identifying major task-related problems within the team. Team members: Become involved in action planning, implement solutions to identify problems	Buller's (1986) problem-solving component subsumes aspects from all of the components described by Beer (1976). Team members practice setting goals, develop interpersonal relations, clarify team roles, and work to

(continued)

Table 1 (continued)

Component	Salas, Rozell, Mullen, & Driskell, 1999	Salas, Priest, & DeRouin, 2005
	and to evaluate those solutions.	improve organizational characteristics through problem-solving tasks. Can have the added benefit of enhancing critical-thinking skills.

has been a reliance on measuring the effectiveness of team-building interventions with process measures. Although implicitly appealing, improvements in processes can not always be linked to improvements in team performance (e.g., Porras & Wilkens, 1980). For example, team performance is typically fashioned by additional environmental and/or organizational characteristics and contingencies that are out of the volitional control of team members. Stated differently, team members are active participants in the enactment of team processes, but must also interact within the larger system to produce more distal performance outputs. Team processes may also be impacted by the larger organizational environment, but not likely to the same degree as team performance outputs.

As a final methodological concern with previous team-building research, there has been an overreliance on subjective indicators of group or organizational performance criteria as dependent measures (Tannenbaum et al., 1992). Though this type of information may be interesting and relevant to measuring participant satisfaction and other affective outcomes that may be impacted by team-building interventions, it is often not concrete enough to allow for accurate predictions of the performance outcomes of team building.

A Recent Advancement in Team-Building Research

Despite the conceptual and methodological issues associated with evaluations of team building, one recent effort has represented advancement over previous reviews by empirically investigating the effectiveness of these interventions. Specifically, Salas and colleagues (1999) responded to a call by Buller (1986) to use the primary focus of the intervention (i.e., goal setting, interpersonal relations, role clarification, and problem solving) as a potential moderating variable. The results of their study failed

to indicate a relationship between the combined set of team-building interventions and team performance ($r = .01$, $k = 16$ effect sizes). Moreover, of the four components of team building, only role clarification proved to be effective, as judged by both objective ($r = .71$) and subjective ($r = .75$) accounts of performance. Interestingly, for objective and subjective measures there was a nonsignificant effect of goal setting ($r = -.06$ and $-.11$, respectively), interpersonal relations ($r = -.38$ and $-.04$, respectively), and problem solving ($r = -.31$ and $.09$, respectively). Also examined in this study were the potential moderating influences of the source of the criterion measurement of performance (i.e., objective vs. subjective), team size, and training duration. For objective measures of performance, there was no evidence of a relationship between team building and performance ($r = -.04$, $k = 8$); for subjective measures, there was a small positive relationship between team building and performance ($r = .14$, $k = 8$). Concerning team size, the results of their study suggested that the effects of team building on performance decreased as a function of the size of the team ($r = -.34$). Finally, there was a slight tendency for the effects of team building to decrease as a function of the duration of the intervention ($r = -.20$).

Taken together, this study enhanced our understanding of the efficacy of team building. However, there were a number of limitations associated with it—limitations that now necessitate the need for additional inquiry. For example, the amount of data analyzed in this study was relatively modest. Specifically, the research findings presented were based on only 16 effect sizes, and thus, it is difficult to discern, with any degree of confidence, the actual effectiveness of these interventions. Moreover, the findings derived from the Salas and colleagues' (1999) study did not necessarily reflect the findings from existing narrative reviews. Finally, their study left many questions unanswered regarding the potential moderating impact of other relevant variables. As an organizing tool provided to summarize the literature in this area, Table 2 provides a summary of five previous investigations into the efficacy of team building, and includes the number of articles reviewed, the years spanned, and other noteworthy features.

In summary the current research was initiated to provide an updated meta-analysis of the team-building literature. Phrased in the form of questions, an examination of these issues will help clarify our understanding of the effectiveness of team building and lead to hypothesized relationships involving the effectiveness of team building, including an investigation of specific moderators. A model that serves to graphically illustrate these hypotheses is presented in Figure 1.

Table 2
Summary of Previous Team-Building Reviews

Review	Study Details	Summary of Findings	Noteworthy Features
Woodman & Sherwood (1980)	30 articles Qualitative data Years spanned: 1964–1978	TB elicits positive affective reactions. The linkage between TB and work group performance remains largely unsubstantiated. TB is more commonly conducted with management teams than groups lower in organizational hierarchies. TB is more commonly conducted with intact and established work teams than new groups. Affective reactions as dependent measures are used more often than objective performance data.	Distinguished between team development and T-group or sensitivity training. Made subjective assessments of the internal validity of the studies reviewed.
DeMeuse & Liebowitz (1981)	36 articles Qualitative data Years spanned: 1962–1980	TB is described as having great promise for improving employee attitudes, perceptions, behaviors, and organizational effectiveness. Eighty-seven percent of the 68 evaluations indicated positive results.[a] Due to the lack of rigorous research designs, firm conclusions concerning the effectiveness of these interventions could not be made.	Coded studies according to research design, sample size, multiple dependent variables, and duration of the TB intervention.
Buller (1986)	9 articles Qualitative data Years spanned: 1964–1981	TB must be more carefully defined. More rigorous experimental designs should be used. There are numerous methodological flaws in previous TB studies. Therefore a clear assessment of the TB and task performance relationship had yet to emerge.	Presented and described a general problem solving approachto TB that added to Beer's (1976) four-component model. Argued that Beer's classification is difficult to use in practice because TB programs usually involve elements from each of the models.
Tannenbaum, Beard, & Salas (1992)	17 articles Qualitative data Years spanned: 1980–1988	The quantity of TB research decreased; however, the quality had improved. Most studies used multiple components in their TB interventions. More researchers began using behavioral and objective measures. Presented evidence to cast doubt the connection between process and performance; TB is effective, but for only perceptions and attitudes.	Presented a comprehensive model of team effectiveness that continues to influence research and theorizing in this field.

(continued)

Table 2 (continued)

Review	Study Details	Summary of Findings	Noteworthy Features
		Post-intervention strategies may be a key mechanism to ensure the long-term effectiveness of these interventions.	
Salas, Rozell, Mullen, & Driskell (1999)	11 articles Quantitative data Years spanned: 1965–1990	No significant effect of TB on performance. A nonsignificant tendency for TB to result in lower performance when measured objectively, but increase performance when subjective measures were used. The role clarification component was more likely to increase performance.[b] The effects of TB decreased as the size of the team increased. The effects of TB decreased as the duration of the intervention increased.	First known attempt to empirically summarize the effectiveness of TB. Assessed a number of moderators, including team building component, team size, training duration, and type of performance measure used (i.e., objective versus subjective).

Note: TB = team building.

a. Many of the 36 studies reported evaluations of multiple dependent variables.

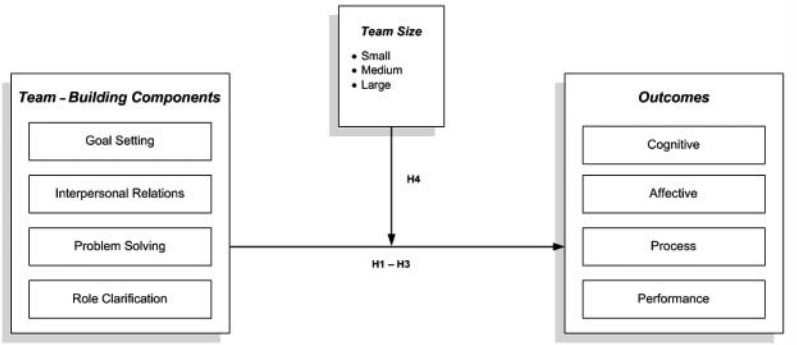
b. This result should be interpreted with caution as it was based on a small number of effect sizes (it's difficult to determine, but likely only three or four effect sizes were used in this calculation).

Hypotheses

Is Team Building Effective?

We agree with Salas and colleagues (1999) that previous narrative reviews have frequently expressed the benefits that can result from team building (e.g., Buller, 1986; DeMeuse & Liebowitz, 1981; Sundstrom, DeMeuse, & Futrell, 1990, Tannenbaum et al., 1992; Woodman & Sherwood, 1980); however, there has been a lack of definitive, compelling evidence concerning the positive effects of team building on team performance. Specifically, previous qualitative reviews of the team-building domain have concluded that evidence of an effect of team building on performance was "inconclusive" (Buller, 1986), "unsubstantiated" (Woodman & Sherwood, 1980), "equivocal" (Tannenbaum et al., 1992), and "mixed" (Sundstrom et al., 1990). Meta-analytic results from one study have suggested there is no overall effect of team building on team performance (Salas et al., 1999). However, there was support for the role-clarification component of team building. Theoretically, one would assume that an intervention focused on

Figure 1
Theoretical Model Depicting Study Hypotheses



improving team functioning would result in positive (as opposed to negative) outcomes. In addition, the moderate support suggested by several narrative reviews (e.g., Tannenbaum et al.) leads us to predict a positive overall effect of team building on team functioning. Thus, we present our first hypothesis:

Hypothesis 1: Team building interventions will result in enhanced team outcomes.

The result from testing this hypothesis will provide for a baseline judgment concerning the efficacy of this particular form of team-development intervention. The omnibus test will therefore allow for an overall assessment of the efficacy of team building, with all independent outcomes from primary studies combined for this analysis.

Is Team Building More Effective for Some Outcomes Than Others?

This framing question asks whether the combined set of team-building interventions is shown to be more useful for improving certain team outcomes than others. Specifically, does team building work better for cognitive outcomes (e.g., declarative knowledge of teamwork competencies), team

member affective outcomes (e.g., trust, team potency), team processes (e.g., coordination, communication), or team performance outcomes (e.g., volume of sales, productivity measures)? This division of team-building outcomes is similar to the commonly discussed cognitive, affective, and skill-based breakdown of general training outcomes (e.g., Kraiger, Ford, & Salas, 1993). However, for the current research, skill-based outcomes are further divided into two additional categories—team processes and more performance-related or productivity-related outcomes.

The division and examination of team-building effectiveness based on specific outcomes is intended to help clarify the often disparate results seen in the literature on these interventions. For example, Woodman and Sherwood's (1980) qualitative review of team building concluded that it was only useful for facilitating affective outcomes, not team performance. Equally pejorative to the position that team building results in improved team performance was the conclusion provided by Buller (1986), who upon reviewing team-building studies conducted through 1980, suggested that the relationship between team building and performance was inconclusive. In yet another review that added to the opaque nature of the efficacy of team building, DeMeuse and Liebowitz (1981) accurately noted that most of the early research on team building relied almost exclusively on perceptual ratings of the dependent variables being studied.

More recent reviewers of the efficacy of team building have reported somewhat different (i.e., more positive) conclusions. For example, the integration of team-building research reported by Salas and colleagues (1999) has served, in many ways, as the focal point for the current research. And, although there was no significant overall effect of team building in their research, there was a small, yet significant, tendency for team building to increase performance when criteria were assessed with subjective measures. In another investigation into the efficacy of team building, Tannenbaum and colleagues (1992) reviewed team-building research conducted in the 1980s and found support for these interventions, especially when the outcome of interest was limited to team member perceptions or attitudes. Finally, Svyantek, Goodman, Benz, and Gard (1999) found that team building positively impacted work-group productivity. Specifically, the largest impact of team building was on productivity measures of cost-effectiveness, with a smaller influence on quantity and quality measures.

In general, previous research that has investigated the effectiveness of team building for improving specific outcomes has been equivocal at best, especially considering performance or productivity measures. At the same time, there has been perhaps the greatest support for the efficacy of team

building for improving affective or attitudinal outcomes. Finally, to our knowledge there have been no reviews citing the correlation between team-building interventions and improvements in cognitive outcomes.

The mixed results found in the literature concerning the effect of team-building interventions on team outcomes lead us to believe that moderators may exist. Therefore, team building impacts certain outcomes more than others. Based on the general characteristics of team building, (e.g., that team building develops interpersonal relations, mutual trust, and open communication between team members), it is likely that team building will have a greater effect on affective outcomes than any on other type of outcomes. As a process intervention, it makes theoretical sense that team building would result in enhanced team member affective outcomes. Finally, although the findings of more distal reviews of team building were taken into consideration, a decision was made to place more credence on recent research, which more often reported positive findings. These findings, along with a complete consideration of the expected benefits of team building, led us to Hypotheses 2a and 2b:

Hypothesis 2a: Team building interventions will result in improved outcomes across each of the four outcome types.

Hypothesis 2b: Team building will be most effective for improving affective outcomes.

Does the Focus of Team Building Moderate Its Effectiveness?

It is certainly important to be able to identify whether team building results in enhanced team functioning, but it is perhaps more informative to know which forms or components of team building are most effective. Advancements in theory in the last quarter century (e.g., Buller, 1986; Salas et al., 1999; Salas, Priest, & DeRouin, 2005) have allowed for the parceling of team-building interventions into four distinct foci (see Table 1). Unfortunately, the single existing empirical integration on the relative efficacy of different components of team building (i.e., Salas et al.) did not provide encouraging results. Combining subjective and objective estimates, there was a slight (nonsignificant) tendency for goal-setting ($r = -.16$), interpersonal relations ($r = -.06$), and problem-solving ($r = -.05$) models of team building to result in decreased performance. It was only the role-clarification component that appeared to be effective for improving performance ($r = .76$). However, upon examination of their data set, it

appears this positive finding for role clarification was supported by the results from only three studies and three effect sizes. Thus, caution and further investigation are warranted before concluding that the role-clarification component of team building is superior to the other forms.

Nonetheless, because the role-clarification component of team building is designed to relieve role stress as created by role ambiguity or role conflict, it is reasonable to anticipate that a significant improvement in team functioning should result. In addition, the role-clarification component of team building emphasizes communication among team members, and thus it is likely that an increase in the level and quality of communication between team members will impact their effectiveness. Unlike the other forms of team building, improvements in role clarity and communication are expected to produce more lasting benefits in terms of team functioning (as assessed through an analysis of the combined set of team outcomes). Combining this theoretical rationale with the preliminary findings of Salas and colleagues (1999) concerning the role-clarification component of team building, it is our belief that a team-building intervention that utilizes a role-clarification focus will provide the most benefit to team functioning. At the same time, it is our view that any carefully thought out team-building intervention should have at least some positive impact on team members. Thus, we also expect that team building that focuses on interpersonal relations, goal setting, or problem solving will also prove useful, at least for the short-term benefits that are typically assessed. Taking these dual considerations into perspective, we present our next set of hypotheses:

Hypothesis 3a: Each of the four components of team building will demonstrate a moderate level of effectiveness for improving team functioning.

Hypothesis 3b: The role clarification component of team building will be most effective for improving team functioning.

Is the Effectiveness of Team Building Moderated by Team Size?

This research will also investigate whether the effectiveness of team building is moderated by the size of the team. At a basic level, "the resources available on a team result from how many people are on it" (Hambrick & D'Aveni, 1992, p. 1449). Sundstrom, McIntyre, Halfhill, and Richards's (2000) review of 83 field studies and experiments conducted with work

groups suggested that the average group size was 11 members. Moreover, Halebian and Finkelstein (1993) have suggested that team size is synonymous with cognitive capability. Providing support for this assertion, Bantel and Jackson (1989) found that larger teams generally have a greater reservoir of cognitive resources than smaller teams. There is little doubt that there are benefits to having medium- to large-sized teams available to perform work tasks rather than smaller teams. However, research has also found that larger teams can facilitate the enactment of other, less desirable, group phenomenon, including the participation leadership effect, the in-group bias effect, the cohesiveness performance effect, and the well-known groupthink effect (cf. Mullen, Anthony, Salas, & Driskell, 1994; Mullen, Brown, & Smith, 1992; Mullen & Copper, 1994; Mullen, Salas, & Driskell, 1989).

From other research, we know that information pooling is critical to team-based decision making. That is, the size of a team may impact the effectiveness with which a team pools common and unique information. For example, Stasser, Vaughan, and Stewart (2000) discussed the tendency for group members to discuss shared information rather than the unique information that is held by team members—an issue that becomes more problematic as teams increase in size. This lack of attention to unique information held by individual team members can lead to mal-informed decisions and occasionally even detrimental results.

Salas and colleagues (1999) presented the only existing research integration that has assessed the relative efficacy of team building for teams of different sizes. They found that the effects of team building on performance decreased as a function of the size of the team, both in objective measures and in subjective measures of performance. These authors concluded that any positive effect of team building is most likely to prevail only in small teams. Similarly, others have argued that as group size increases, members' liking for the group (Indik, 1965) and performance (Mullen, 1987) tend to decrease. Taking these findings into consideration, the current research examined team size as a moderator of the effectiveness of team-building efforts on team performance. In this study, we seek to replicate Salas and colleagues' findings. However, rather than simply correlating the database of effect sizes with their associated team sizes, the current research will examine the efficacy of team building for three distinct subgroup classifications of team size: small teams (i.e., less than 5 members); medium-sized teams (i.e., 5 to 10 members); and large teams (i.e., greater than 10 members). In addition, although there is some reason to expect that larger teams have an increased cognitive capacity with which to perform tasks (e.g., Bantel & Jackson, 1989), the other negative issues often associated with increased group size are expected to be the overwhelming influences on the performance of teams. Therefore, it is reasonable to suggest that larger teams

are already performing at a lower level than medium or small teams, and would therefore exhibit enhanced benefits from team-building interventions. Stated differently, as teams increase in size, it is more likely they will show substantial benefits from team building. This expectation, although supported by theory, is in direct contradiction to the findings reported by Salas and colleagues. The following hypothesis is proposed to investigate this assertion.

Hypothesis 4: Large teams will show greater benefits from team building than small- or medium-sized teams.

Method

Literature Search

A comprehensive literature review was conducted to identify published and unpublished studies relevant to the effects of team building on the four team outcomes. As a starting point for our literature review, we conducted an online search of the Defense Technical Information Center (DTIC), Google Scholar ©, and the most common academic search databases (e.g., PsycINFO). The particular keywords that were used included, but were not limited to, the following terms: team building, team development, team goal setting, interpersonal relations, problem solving, role clarification, group building, and group development. In addition to these search techniques, key articles relevant to team building were inspected by hand for additional, potentially useful primary studies (e.g., Buller, 1986; DeMeuse & Liebowitz, 1981; Salas et al., 1999; Tannenbaum et al., 1992; Woodman & Sherwood, 1980). This ancestry approach was later extended to each of the articles that were included in the database in an effort to ensure that no fugitive studies had been overlooked. In the end, the literature search process resulted in 103 articles being identified for potential inclusion in the database.

Criteria for inclusion. Specific limits were placed on the search to include articles published from 1950 to 2007. This date range was selected because previous reviews had not uncovered any published or unpublished team-building evaluations prior to 1950. Moreover, studies were included only if they involved adult, human, and nonclinical populations, and focused on teams and not groups. Finally, studies had to report data sufficient to calculate an effect size (r) assessing the relationship between team-building interventions and outcomes.

Coding Procedure

Following the literature review, three authors independently coded each of the relevant studies on 14 categories: (a) nature of the organization and participant sample; (b) team type; (c) number of teams; (d) average team size; (e) predictor reliability; (f) level of analysis of predictor; (g) criterion reliability; (h) level of analysis of the criterion; (i) criterion report type(s); (j) criterion description(s); (k) focus/component of team building (i.e., goal setting, interpersonal relations, problem solving, role clarification); (l) effect size(s); (m) study design type; and (n) recommendation for inclusion. Concerning the level of analysis coding, the initial database consisted of 69 effect sizes—60 at the team level and 9 at the individual level. However, mixing levels of analysis is not recommended for research integrations for numerous reasons (e.g., Gully, Devine, & Whitney, 1995; Hunter & Schmidt, 1990). Therefore, only the team-level outcomes were analyzed in the current research, reducing our database to the 60 team-level correlations.

Another point concerning the coding process that bears clarification is the assessment of the focus or component of team building. It was expected that many of the interventions being examined would consist of multiple components of team building. Thus, coders were instructed to allocate a total of 100 percentage points to each of the four components of team building. For example, if a coder believed an intervention consisted of equal parts role clarification and interpersonal relations, each was coded with 50 percentage points, whereas problem solving and goal setting were given a 0 for this coding category. For this process, the coders were instructed to closely examine the effort, intensity, and primary focus of the intervention. They were instructed further that they would need to justify the weights they assigned during a consensus discussion of the final articles included in the database. In this discussion, the three coders came to a consensus concerning the allocation of percentage points to each team-building intervention.

Rater reliability. Articles were coded by one of three of the authors. To aid in the consistency of coding, the three authors met early in the process to discuss the 14 pieces of information that were to be extracted from each of the 103 articles. Moreover, the coders had previously pilot tested the coding scheme, and possessed satisfactory level of expertise in the substantive areas being coded. To estimate the reliability with which the coders were evaluating the primary studies, three authors independently coded the same set of 20 articles (the reliability sample, representing approximately 19% of the total 103 articles selected for coding) to determine interrater reliability

by calculating intraclass correlation coefficients (ICCs; Nunnally, 1978; Shrout & Fleiss, 1979). The coders then each independently coded approximately one third of the remaining articles.

During the coding process, the decision to include or exclude the article for the final analyses was of primary concern. Upon inspection of the 20 studies coded by each of the three coders, it was revealed that all three coders were in agreement 85% of the time with their global evaluation to include (or not to include) particular studies. The ICC for this assessment was satisfactory (3, k) = .85. Importantly, for the three studies for which there was not complete initial agreement the authors met to discuss the discrepancies and a consensus was reached. The ICCs calculated for seven of the categories ultimately used in either main or exploratory analyses ranged from ICC (3, k) = .85 to ICC (3, k) = 1.0. In addition, the agreement among the three coders across all seven categories resulted in an ICC (3, k) = .96. The remaining six categories that were utilized for the ICC calculations included: team type, number of teams in the primary studies under investigation, average team size, focus/component of the team-building intervention, effect size estimate, and study design. The ICCs for these categories were .94, .94, .97, .94, .91, and 1.0, respectively. In conclusion, the consistency and agreement among the raters was very good.

Meta-Analysis Procedure

The software for the Hunter-Schmidt meta-analysis methods was used to analyze the data (Schmidt & Le, 2005). This program provides output that includes, but is not limited to: (a) the mean true score correlation, (b) the variance and standard deviation of true score correlations, (c) credibility intervals, (d) estimates of the variance and standard deviation in observed correlations due to artifacts, and (e) the percentage of variance attributable to observed correlations after the removal of artifacts. In addition, the software employs a random effects model to combine effect sizes from primary studies. Random effects meta-analysis models allow the true effect sizes to vary, in contrast with fixed-effects models that assume the true effect sizes have fixed values. Moreover, the random effects model is considered to be "more realistic than the fixed-effect model on the majority of occasions" (Field, 2001, p. 162).

Effect-Size Calculations

Before calculating a meta-analytic estimate of the relationship between variables, effect sizes culled from primary studies must be prepared for

entry into the database. During this process, effect sizes from primary studies were converted to the common metric *r* (correlation). Thus, when necessary, primary study effect sizes reported as other statistics (e.g., *t*, *F*, *d*, χ^2, or *Z*) were transformed using the formulas found in Hunter and Schmidt (2004). Once placed on this common metric of effect size, the results of independent tests found in primary studies can be combined and assessed for fit with predicted hypotheses in the meta-analysis. However, it is common knowledge that when a study contains multiple effect sizes, they are stochastically dependent (Shadish, Cook, & Campbell, 2002). These dependencies violate the statistical assumption of independent effect sizes. The recommended solution is to average or combine effect sizes from different measures of the same sample within a single study prior to combining results from multiple studies.

Corrections for Unreliability and Range Restriction

It is common in meta-analyses to make attempts to correct obtained reliability coefficients for measures of the predictor, criterion, or both (e.g., Hunter & Schmidt, 1990; Johnson, Mullen, & Salas, 1995). Unfortunately, original studies often fail to report all of the auxiliary information necessary to perform corrections for study artifacts. Such was the case with the current database. As a result, artifact distribution meta-analysis was employed, rather than correcting each effect size individually for unreliability.

Authors of meta-analytic integrations also make attempts to correct for the effects of direct or indirect range restriction. Performing corrections for range restriction will generally result in a combined estimate that is more accurate than had no corrections been performed at all (e.g., Hunter, Schmidt, & Le, 2006). However, in the current investigation, primary studies did not report information on restricted and unrestricted samples that would be necessary to perform corrections for range restriction. We therefore cannot correct for range restriction in the current meta-analysis. Consequently, our effect-size estimates may be conservative, in the sense that they may underestimate the true effect of team building if the biasing effects of range restrictions could be corrected.

Weighting

Primary study effect sizes included in the current series of meta-analytic integrations were weighted by their sample sizes. It is presumed that effect sizes obtained from studies with larger sample sizes are more stable (i.e., accurate) than those culled from studies with small sample sizes. Although

it is important to remember that large sample studies are not necessarily more valid, recent simulation studies have determined that a very accurate combined effect size can be obtained through procedures used to weight primary studies by sample size (e.g., Field, 2005).

Results

Description of the Database

A total of 60 correlations were obtained from 20 studies. These 60 effect sizes represented 1,562 teams, with a median team size of approximately 9 members. Although these 60 effect sizes were not all from independent samples, every subgroup analysis that was performed included only independent samples. Otherwise, for the overall assessment of the influence of team building on the combined set of team outcomes, there were 26 independent samples (see Table 3). Of the studies included in the meta-analytic database, 14 were published and 6 were unpublished. Table 3 provides a description of the key information derived from each primary study.

Publication/Availability Bias Detection

When performing any meta-analysis it is difficult to determine whether all relevant studies have been located. Existing empirical evaluations that are elusive to locate and retrieve are commonly referred to as fugitive literature (Rosenthal, 1994). To the extent that this fugitive literature represents a substantial proportion of the conducted evaluations in any one area, there is a possibility of publication or availability bias. The file drawer problem in meta-analyses presents itself when there is a concern that the studies that find their way into publication are simply the ones that show significant results, whereas evaluations with nonsignificant findings are relegated to file drawers (Rosenthal, 1979). To alleviate this concern, Rosenthal's (1979) file drawer analysis can be used to provide an estimate of the number of unpublished studies relegated to file drawers. This analysis designates the number of these articles, which average null results, that would be required to bring the significance level for a set of studies down to the just-significant level (Hunter & Schmidt, 2004). However, it has been demonstrated that the widely used fail-safe file drawer analysis is essentially irrelevant and can result in incorrect estimates of the size of the file drawer (Scargle, 2000). Instead, funnel plots can serve as an alternative to file drawer analysis for the purpose of detecting the possibility of bias.

Table 3
Meta-Analytic Database

Study Author(s)	Year	Design Type	Effect Size, *r*	GS	IR	PS	RC	Team Type	Number of Teams	Average Team Size	Criterion Description
Boss & McConkie	1981	SGPP	0.91	0.00	1.00	0.00	0.00	Executive/ Management	2	10.00	Affective
Boss & McConkie	1981	SGPP	0.85	0.00	1.00	0.00	0.00	Executive/ management	2	10.00	Process
Boss & McConkie	1981	SGPP	0.89	0.00	1.00	0.00	0.00	Executive/ management	2	10.00	Performance
Bragg & Andrews	1973	PPWC	0.83	0.00	0.00	1.00	0.00	Production	3	32.00	Performance
Buller & Bell	1986	SGPP	0.86	0.00	0.00	1.00	0.00	Production	12	16.00	Process
Buller & Bell	1986	SGPP	0.85	1.00	0.00	0.00	0.00	Production	8	16.00	Process
Buller & Bell	1986	SGPP	0.70	0.50	0.00	0.50	0.00	Production	12	16.00	Process
Buller & Bell	1986	SGPP	0.59	1.00	0.00	0.00	0.00	Production	8	16.00	Performance
Buller & Bell	1986	SGPP	0.57	0.00	0.00	1.00	0.00	Production	12	16.00	Performance
Buller & Bell	1986	SGPP	0.66	0.50	0.00	0.50	0.00	Production	12	16.00	Performance
Bushe & Coetzer	1995	PPWC	0.40	0.00	0.00	1.00	0.00	Project	16	4.00	Affective
Bushe & Coetzer	1995	PPWC	0.36	0.50	0.00	0.00	0.50	Project	16	4.00	Affective
Bushe & Coetzer	1995	PPWC	0.28	0.00	0.00	1.00	0.00	Project	16	4.00	Process
Bushe & Coetzer	1995	PPWC	0.33	0.50	0.00	0.00	0.50	Project	16	4.00	Process
Bushe & Coetzer	1995	PPWC	0.92	0.00	0.00	1.00	0.00	Project	16	4.00	Performance
Bushe & Coetzer	1995	PPWC	0.95	0.50	0.00	0.00	0.50	Project	16	4.00	Performance
Cohen	1993	SGPP	0.16	0.00	1.00	0.00	0.00	Action/ performing	4	16.67	Affective
Cohen	1993	PPWC	0.17	0.00	1.00	0.00	0.00	Action/ performing	4	16.67	Affective
Cohen	1993	SGPP	0.13	0.00	1.00	0.00	0.00	Action/ performing	4	16.67	Process

(continued)

Table 3 (continued)

Study Author(s)	Year	Design Type	Effect Size, *r*	GS	IR	PS	RC	Team Type	Number of Teams	Average Team Size	Criterion Description
Cohen	1993	PPWC	0.18	0.00	1.00	0.00	0.00	Action/ performing	4	16.67	Process
Dionne	1998	SGPP	0.20	0.00	1.00	0.00	0.00	Project	54	4.50	Process
Dionne	1998	SGPP	0.05	0.00	1.00	0.00	0.00	Project	54	4.50	Performance
Eden	1986	PPWC	0.13	0.33	0.33	0.00	0.33	Action/ performing	16	30.00	Performance
Eden	1986	PPWC	0.36	0.33	0.33	0.00	0.33	Action/ performing	16	30.00	Process
Friedlander	1967	PPWC	0.00	0.00	0.50	0.50	0.00	Project	12	10.00	Affective
Friedlander	1967	PPWC	0.16	0.00	0.50	0.50	0.00	Project	12	10.00	Process
Friedlander	1967	PPWC	0.21	0.00	0.50	0.50	0.00	Project	12	10.00	Performance
Gibson	2001	PPWC	0.24	1.00	0.00	0.00	0.00	Service	71	5.00	Affective
Gibson	2001	PPWC	–0.12	1.00	0.00	0.00	0.00	Service	71	5.00	Performance
Howard	1979	PPWC	0.25	0.00	1.00	0.00	0.00	Service	3	9.67	Affective
Howard	1979	PPWC	0.40	0.00	1.00	0.00	0.00	Service	3	9.67	Performance
Huang, Wei, Watson, & Tan	2002	PPWC	0.97	1.00	0.00	0.00	0.00	Production	48	5.00	Affective
Huang, Wei, Watson, & Tan	2002	PPWC	0.97	1.00	0.00	0.00	0.00	Production	48	5.00	Process
Huang, Wei, Watson, & Tan	2002	PPWC	0.68	1.00	0.00	0.00	0.00	Production	48	5.00	Performance
Hughes, Rosenbach, & Clover	1983	SGPP	0.50	0.50	0.50	0.00	0.00	Project	2	74.00	Affective
Hughes, Rosenbach, & Clover	1983	PPWC	0.23	0.50	0.50	0.00	0.00	Project	2	68.00	Affective

(continued)

Table 3 (continued)

Study Author(s)	Year	Design Type	Effect Size, *r*	GS	IR	PS	RC	Team Type	Number of Teams	Average Team Size	Criterion Description
Hughes, Rosenbach, & Clover	1983	SGPP	0.52	0.50	0.50	0.00	0.00	Project	2	74.00	Cognitive
Hughes, Rosenbach, & Clover	1983	PPWC	0.03	0.50	0.50	0.00	0.00	Project	2	68.00	Cognitive
Kimberley & Nielsen	1975	SGPP	0.48	0.00	0.00	1.00	0.00	Production	180	10.00	Affective
Kimberley & Nielsen	1975	SGPP	0.38	0.00	0.00	1.00	0.00	Production	180	10.00	Process
Kimberley & Nielsen	1975	SGPP	0.14	0.00	0.00	1.00	0.00	Production	90	10.00	Performance
Kimberley & Nielsen	1975	SGPP	0.19	0.00	0.00	1.00	0.00	Production	90	10.00	Performance
Longenecker, Scazzero, & Stansfield	1994	SGPP	0.83	1.00	0.00	0.00	0.00	Production	2	45.00	Performance
Miller	1997	SGPP	0.18	0.10	0.10	0.70	0.10	Action/ performing	16	7.38	Affective
Miller	1997	SGPP	0.06	0.10	0.10	0.70	0.10	Action/ performing	16	7.38	Process
Mitchell	1986	SGPP	0.72	0.00	1.00	0.00	0.00	Project	12	4.50	Affective
Mitchell	1986	SGPP	0.25	0.00	1.00	0.00	0.00	Project	12	4.50	Affective
Mitchell	1986	PPWC	0.78	0.00	1.00	0.00	0.00	Project	9	4.50	Affective
Morrison & Sturges	1980	SGPP	0.67	0.00	.50	0.00	0.50	Executive/ management	2	12.00	Affective
Morrison & Sturges	1980	SGPP	0.47	0.00	0.50	0.00	0.50	Executive/ management	2	12.00	Process

(continued)

Table 3 (continued)

Study Author(s)	Year	Design Type	Effect Size, *r*	GS	IR	PS	RC	Team Type	Number of Teams	Average Team Size	Criterion Description
Morrison & Sturges	1980	SGPP	0.47	0.00	0.50	0.00	0.50	Executive/ management	2	12.00	Performance
Wegenast	1983	SGPP	0.39	0.00	0.25	0.50	0.25	Service	4	6.00	Affective
Wexler	1990	SGPP	0.24	0.00	0.00	1.00	0.00	Service	4	6.00	Process
Wexler	1990	SGPP	–0.22	0.00	1.00	0.00	0.00	Service	4	4.50	Process
Wexler	1990	PPWC	0.39	0.00	0.00	1.00	0.00	Service	4	5.50	Process
Wexler	1990	PPWC	–0.46	0.00	1.00	0.00	0.00	Service	4	4.75	Process
Woodman & Sherwood	1980	PPWC	0.10	0.50	0.00	0.50	0.00	Project	67	3.50	Cognitive
Woodman & Sherwood	1980	PPWC	0.10	0.50	0.00	0.50	0.00	Project	67	3.50	Affective
Woodman & Sherwood	1980	PPWC	0.23	0.50	0.00	0.50	0.00	Project	67	3.50	Process
Woodman & Sherwood	1980	PPWC	0.15	0.50	0.00	0.50	0.00	Project	67	3.50	Performance

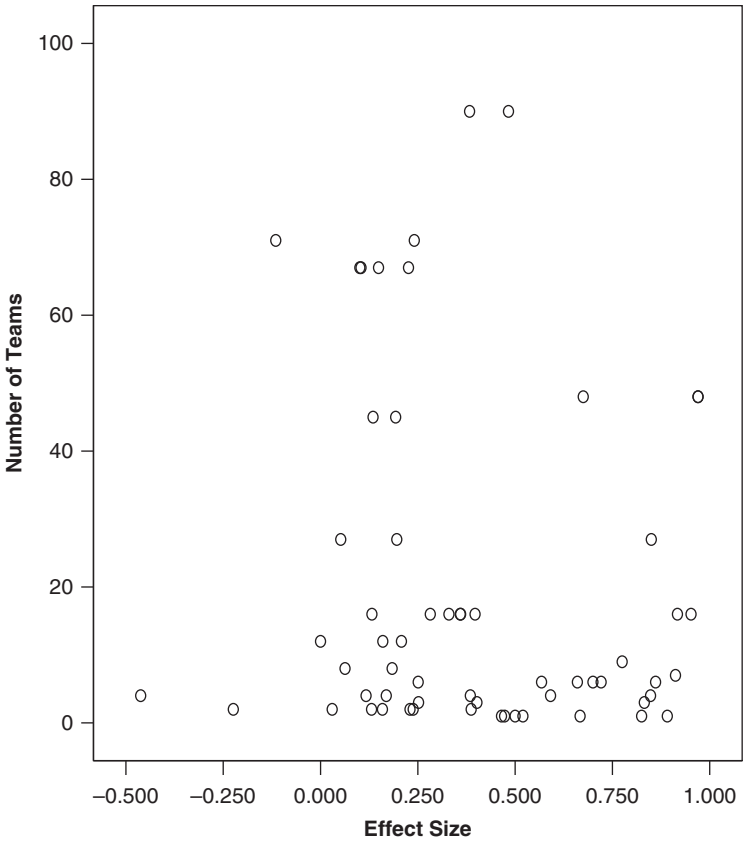
Note: GS = goal-setting team building; IR = interpersonal relations team building; PS = problem-solving team building; RC = role-clarification team building; PPWC = pre–post with control group comparison; SGPP = single-group pre–post comparison.

Funnel plot. The funnel plot (Light & Pillemer, 1984) is a simple visual tool (i.e., scatterplot) for detecting the presence of publication or other availability bias in meta-analysis. To apply this technique, effect sizes are graphed on the horizontal axis, whereas study sample sizes (*N*) are graphed on the vertical axis. The idea is that, in the absence of bias, the results of small sample studies will scatter widely at the bottom of the graph, with the spread narrowing for larger *N* studies; thus taking the form of an inverted funnel (Hunter & Schmidt, 2004). If publication bias is present in the data, the figure will be asymmetrical and often truncated in the lower left-hand portion of the scatterplot. As seen in Figure 2, the data from primary studies included in the current meta-analysis form a somewhat regular funnel. Thus, it appears that the small sample studies are spread rather evenly across the effect-size continuum, making the possibility of publication or availability bias less of a concern in this research.

Meta-Analytic Results

The meta-analytic results for the hypotheses investigated in this research are presented in Tables 4 through 6. These tables show a number of pieces of information, including: the number of teams in each analysis (*N*); the number of independent effect sizes (correlations) in each analysis (*k*); the mean weighted observed correlation ($\bar{r}$); the 80% confidence interval for that correlation; the estimated true score correlation (ρ); the standard deviation of this true score correlation (SD_ρ); the 80% credibility interval (10% CV and 90% CV); and the percentage of variance accounted for by statistical artifacts. Confidence and credibility intervals are useful as aids in providing the best estimate of the true nature of the relationships between two variables (Whitener, 1990). Confidence intervals are applied to observed scores, center on a single mean score, and reflect the effects of sampling error. Credibility intervals, on the other hand, are particularly meaningful because they take into account information about the distribution of effect sizes after other research artifacts have been taken out. Credibility intervals can also be useful for determining whether moderators are operating (Whitener, 1990). Similarly, the percentage of variance estimate provides information concerning the variance in observed correlations that is the result of statistical artifacts. In general, the higher the percentage, the more certain we can be that additional moderators are not operating. The following sections more closely examine the results for the individual hypotheses that were put forward in this research.

Figure 2
Funnel Plot for Detecting the Possibility of Publication or Availability Bias



Team-building effectiveness. The omnibus test representing all team-building interventions and outcomes resulted in a significant tendency for these interventions to improve team outcomes. Specifically, the mean true score correlation of .31 represents a moderate effect (10% CV = .11; 90% CV = .52) and provided support for Hypothesis 1. This analysis included 26 independent effect sizes and was based on a total sample size of 579 teams.

Table 4
Analysis of the Effectiveness of Team Building Based Upon Outcome Type

Outcome Type	N	k	$\bar{r}$	CI_r 10%	CI_r 90%	ρ	SD_ρ[b]	10% CV	90% CV	% Var. Acct.[c]
Cognitive	71	3	.11	.06	.17	.13	.00	.13	.13	1,394.24
Affective	482	19	.41	.34	.49	.44	.18	.21	.66	55.42
Process	485	20	.39	.31	.46	.44	.20	.18	.69	54.04
Performance	524	18	.25	.16	.33	.26	.23	−.03	.55	45.11
All outcomes	579	26	.28	.22	.29	.31	.16	.11	.52	69.35
All outcomes[a]	1,562	60	.34	.29	.38	.37	.21	.10	.64	50.64

Note: k = number of correlations coefficients on which each distribution was based; $\bar{r}$ = mean observed correlation; CI_r 10% = lower bound of the confidence interval for observed r; CI_r 90% = upper bound of the confidence interval for observed r; ρ = estimated true correlation between the predictor construct and the relevant criterion (fully corrected for measurement error in both the predictor and the criterion); SD_ρ = estimated standard deviation of the true correlation; 10%CV = lower bound of the credibility interval for each distribution; 90% CV = upper bound of the credibility interval for each distribution; % var. acct. = percentage of observed variance accounted for by statistical artifacts.

a. This result represents the entire database and patently violates the assumption of independent effect sizes as many studies contributed more than one effect size. It is shown here for illustrative purposes only.

b. The SD_ρ being zero indicates that the real variance of the true correlation is zero—there is only one value of the true correlation underlying all the studies. It is a consequence of the percentage of variance accounted for estimate being greater than 100%. This result also indicates that there should be no additional moderators operating for this analysis.

c. The percentage of variance explained estimate being greater than the theoretical maximum value of 100% indicates that sampling error and other study artifacts explain all of the observed variation in the effect sizes (correlations) across studies. The estimated value is greater than 100% because of second-order sampling error.

Here, cognitive, affective, process, and performance outcomes were averaged for each primary study before figuring into the combined estimate. The only exception was for six of the primary studies that reported the results of team-building interventions for separate samples within the same study. For these instances only, there were multiple correlations from single studies that contributed to the combined estimate. However, care was taken to ensure the assumption of independent effect sizes was not violated.

Investigation of separate outcomes. From the overall database of 60 effect sizes, this research also assessed the impact of team building on distinct

Table 5
Analysis of the Effectiveness of Team Building Based Upon Team-Building Component

Team-Building Component	*N*	*k*	$\bar{r}$	CI_r 10%	CI_r 90%	ρ	SD_ρ[a]	10% CV	90% CV	% Var. Acct.[b]
Goal setting	258	10	.34	.21	.47	.37	.27	.02	.71	37.65
Interpersonal relations	140	13	.23	.15	.31	.26	.00	.26	.26	233.50
Problem solving	326	11	.23	.16	.29	.24	.00	.24	.24	114.27
Role clarification	54	5	.32	.22	.42	.35	.00	.35	.35	322.06

Note. *k* = number of correlations coefficients on which each distribution was based; $\bar{r}$ = mean observed correlation; CI_r 10% = lower bound of the confidence interval for observed *r*; CI_r 90% = upper bound of the confidence interval for observed *r*; ρ = estimated true correlation between the predictor construct and the relevant criterion (fully corrected for measurement error in both the predictor and the criterion); SD_ρ = estimated standard deviation of the true correlation; 10% CV = lower bound of the credibility interval for each distribution; 90% CV = upper bound of the credibility interval for each distribution; % var. acct. = percentage of observed variance accounted for by statistical artifacts.

a. The SD_ρ being zero indicates that the real variance of the true correlation is zero—there is only one value of the true correlation underlying all the studies. It is a consequence of the percentage of variance accounted for estimate being greater than 100%. This result also indicates that there should be no additional moderators operating for this analysis.

b. The percentage of variance explained estimate being greater than the theoretical maximum value of 100% indicates that sampling error and other study artifacts explain all of the observed variation in the effect sizes (correlations) across studies. The estimated value is greater than 100% because of second-order sampling error.

outcomes. Specifically, we evaluated the impact team building had on cognitive, affective, process, and performance outcomes. The number of effect sizes for these subgroup analyses ranged from 3 to 20. Table 4 presents the results of these investigations.

In evaluating the impact of team building on cognitive outcomes, three effect sizes were analyzed ($N = 71$ teams). The estimated true score correlation (ρ) for the relationship between team building and improvements in cognitive outcomes was .13. Although it appears that team building has a negligible impact on these outcomes, this result should be interpreted with extreme caution since there were only three effect sizes available to combine. For the analysis of the influence of team building on affective outcomes, the estimated true score correlation was .44 (10% CV = .21; 90% CV = .66). This result represented the accumulation of data from 19 effect sizes and a total of 482 teams. Concerning process outcomes, the results

Table 6
Analysis of the Effectiveness of Team Building Based Upon Team Size

Team Size	*N*	*k*	$\bar{r}$	CI_r 10%	CI_r 90%	ρ	SD_ρ[a]	10% CV	90% CV	% Var. Acct.[b]
Small	178	7	.26	.15	.37	.28	.12	.12	.44	74.22
Medium	340	10	.25	.14	.36	.27	.21	.00	.54	45.72
Large	61	9	.54	.44	.64	.66	.00	.66	.66	197.54

Note: k = number of correlations coefficients on which each distribution was based; $\bar{r}$ = mean observed correlation; CI_r 10% = lower bound of the confidence interval for observed r; CI_r 90% = upper bound of the confidence interval for observed r; ρ = estimated true correlation between the predictor construct and the relevant criterion (fully corrected for measurement error in both the predictor and criterion); SD_ρ = estimated standard deviation of the true correlation; 10% CV = lower bound of the credibility interval for each distribution; 90% CV = upper bound of the credibility interval for each distribution; % var. acct. = percentage of observed variance accounted for by statistical artifacts.

a. The SD_ρ being zero indicates that the real variance of the true correlation is zero—there is only one value of the true correlation underlying all the studies. It is a consequence of the percentage of variance accounted for estimate being greater than 100%. This result also indicates that there should be no additional moderators operating for this analysis.

b. The percentage of variance explained estimate being greater than the theoretical maximum value of 100% indicates that sampling error and other study artifacts explain all of the observed variation in the effect sizes (correlations) across studies. The estimated value is greater than 100% because of second-order sampling error.

suggested an estimated true score correlation of .44 (k = 20 correlations; N = 485 teams). The 80% credibility interval for this finding ranged from .18 to .69. Finally, for the analysis on performance outcomes, the results indicated an estimated true score correlation of .26 (k = 18; N = 52).

All told, the results of the analyses focused on the influence of team building on separate outcomes provided moderate support for Hypothesis 2a. Specifically, team building appeared to be effective for improving each of the four outcomes. However, it should be pointed out that the credibility interval for performance outcomes barely included 0 (10% CV = –.03; 90% CV = .55). At the same time, Hypothesis 2b received only partial support. Although it was posited that team building would be most effective for improving affective outcomes, the results from this study suggested that team building was slightly more effective for improving process outcomes (ρ = .439 vs. .437). However, given the considerable overlap in both confidence and credibility intervals for these analyses (see Table 4), it is difficult to determine with any certainty which outcome type is most greatly affected by team-building interventions. Moreover, team building did appear to be

more effective for improving affective outcomes than either cognitive or performance outcomes (ρs = .13 and .26, respectively).

Focus of team building. A total of 39 correlations were meta-analyzed to assess the differential impact of various team-building components. For this analysis, the interventions described in primary studies were closely scrutinized to determine which component(s) of team building was the major focus. In some instances it was determined that the intervention had a multiple focus on two or more separate components. Therefore, in accumulating the study results, correlations from primary studies were sometimes included in multiple categories for the purposes of analysis so that there was some degree of overlap regarding the analysis of team-building components. Table 5 presents the results of these analyses.

Ten effect sizes, representing 258 teams, were meta-analyzed for the investigation into the efficacy of the goal-setting component of team building. The estimated true score correlation was .37, indicating a moderate effect on the combined set of team outcomes. The remaining team-building components, namely, interpersonal relations, problem solving, and role clarification, also resulted in moderate effect sizes. The estimated true mean score correlations were .26 (k = 13, N = 140), .24 (k = 11, N = 326), and .35 (k = 5, N = 54), respectively. The credibility intervals for these analyses were also calculated (see Table 5). All calculated intervals of the true mean score correlation did not include 0 (the 10% CV > .00 for all cases), thus indicating that the true mean score correlations are positive in most of their populations. In other words, the effects of team building based on all the components examined generalize across most situations and settings. The results of this analysis confirmed Hypothesis 3a, as each of the team-building components resulted in improved team functioning. However, there was only partial support for Hypothesis 3b. Specifically, though the role-clarification component of team building (ρ = .35) appeared to be superior to either the interpersonal relations (ρ = .26) or problem-solving (ρ = .24) components, the goal-setting component (ρ = .37) appeared to work best of all.

Team size. Twenty-six effect sizes were used to assess the potential moderating influence of team size on the ability of team building to improve team functioning. The results for this analysis are presented in Table 6. Effect sizes were categorized into three groups based on the average team size. Effect sizes grouped into the small team category included those that averaged less than 5 team members; the medium size category included

those effect sizes based on an average team size of 5 to 10 members; and the large team category included those effect sizes based on an average team size of more than 10 members.

For small teams, a meta-analysis of 7 effect sizes, representing 178 teams, indicated a mean true score correlation of .28. For medium-sized teams of 5 to 10 members, an analysis of 10 effect sizes representing 340 teams resulted in a mean true score correlation of .27. Thus, for both small- and medium-sized teams there was a moderate effect of team building on team functioning. However, the results of a meta-analysis of team building for large teams, which consisted of 9 effect sizes ($N = 61$), suggested that the greatest impact of team building was upon teams that are large in size ($\rho = .66$). This rather large effect is in direct support of Hypothesis 4, which posited that the influence of team building would be most profound in large teams, as compared to small- or medium-sized teams.

Discussion

The present study was conducted to answer the question of whether team building works. The results are encouraging—they are suggestive of the idea that team building does improve team outcomes. Specifically, process and affective outcomes were most improved by team-building interventions. Moreover, all the components (i.e., role clarification, goal setting, interpersonal relations, and problem solving) of team building had a moderate effect on outcomes but the goal-setting and role-clarification components had the largest effect. Although teams of all sizes benefited from team building, large teams appeared to benefit the most.

Despite the use of sophisticated meta-analytic techniques that were employed to assist in the examination of the four hypotheses that served as the focus for this research, our ability to definitively address the framing questions put forth and test for the hypothesized effects was dependent on the research available to us. In some instances, small subgroup effect sizes precluded fine-tuned assessment and the derivation of definitive conclusions. However, the findings from the current research were still instructive in many ways. As noted, the results from this study suggest that we are beginning to find some positive empirical support for the effectiveness of these commonly applied team-development strategies.

What else have we learned? In addition to finding support for the effectiveness of team building in general, it was deemed equally important to understand whether the effectiveness of team building was moderated by

the specific outcomes that are targeted. Our second hypothesis posited that team-building interventions would result in improved outcomes across each of the four outcome types. This hypothesis was supported, with team building shown to be effective for each of the outcome types examined. Of the four outcomes of interest, the results for cognitive outcomes were found to be the least robust. However, given the exceptionally small number of correlation coefficients that contributed to this subgroup analysis (i.e., $k = 3$), this finding may be properly viewed with considerable skepticism. The second part of this hypothesis was based upon prior theory and research in this area, and posited that team building would be most effective for improving affective outcomes. The findings indicated that team building is indeed highly effective for improving team member affective outcomes, but may be just as useful for facilitating improvements in team processes. Thus, it remains an open question as to which outcome is most greatly affected by team building—team processes or team member affective outcomes.

Our next research inquiry led us to investigate whether the focus of team building moderated its effectiveness. First, it was found that each of the team-building components were useful for enhancing team functioning, with the estimated true score correlations ranging from .24 to .37. Similar to the Salas and colleagues' (1999) meta-analysis, our data suggested that the role clarification component had the most impact on team outcomes. However, in contrast to Salas and colleagues, it appeared that the goal-setting component of team building was equally and perhaps more effective for improving team functioning as the role-clarification component. The results from this research also suggested that the interpersonal relations and problem-solving components were less effective.

This research also examined whether team building is more effective for larger teams than smaller teams. Specifically, it was hypothesized that the effect sizes for small- and medium-sized teams would be weaker than those observed from team-building interventions conducted with large teams. Our prediction was supported—there was a substantial estimated true score correlation found for large teams and more modest results with teams that were classified as small or medium. One possible explanation for this result is the likely pre-intervention state of teams of various sizes. Although larger teams generally have a greater reservoir of available cognitive resources (Bantel & Jackson, 1989), it is also suggested that they are plagued by problems such as groupthink, the participation leadership effect, and the cohesiveness performance effect (e.g., Mullen et al., 1989). Given all the potential problems that can accompany larger teams, it is feasible that larger teams begin in a more negative state than smaller teams. Thus, there

may be less room for improvements from team building to manifest themselves with teams of smaller size, which are typically more cohesive and less likely to experience these problems.

In conclusion, this research extends the team-development literature by providing an empirical assessment of the relationship between team building interventions and outcomes. In addition, a number of potential moderators were also investigated. By extending the meta-analysis by Salas and colleagues (1999), more direction can be given to researchers who seek to further investigate specific questions regarding the effectiveness of team building. However, the fact that our findings were not that consistent with those provided by Salas and colleagues deserves a brief explanation. Specifically, how should we interpret the present findings in light of the contrasting findings of Salas and colleagues?

Importantly, our database included approximately twice as many articles (20 vs. 11) and nearly four times as many correlations (60 vs. 16) as the previous effort. In addition, although their study was published in 1999, the last included article in their database was published in 1990. In contrast, our database included seven sources put forward since 1990. The increased number of effect sizes included in our analyses should have resulted in more stable estimates of the relationships under investigation. It is also important to point out that 4 of the 11 articles included in the research by Salas and colleagues were excluded from our analyses. In one instance, we chose to include a published version of a dissertation included in the previous integration. In two other cases, we determined there was no real team-level intervention or team-level outcomes under investigation. Finally, for a fourth article, only individual-level outcomes were reported, making that study unusable for our purposes. These differences in meta-analytic databases, combined with the reality of using slightly more conservative meta-analysis procedures, may have led to the divergent findings between the two studies. Specifically, the current research used meta-analysis methods (i.e., Hunter & Schmidt, 1990; Hunter, Schmidt, & Jackson, 1982) that can be considered more conservative (Johnson et al., 1995) than the methods employed by Salas and colleagues (1999), whose data analyses were based on the methods employed by Rosenthal and colleagues (i.e., Rosenthal, 1991; Rosenthal & Rubin, 1988).

Implications for Research and Practice

What should academicians and human resources practitioners do when investigating or recommending team-building interventions? The answers

seem to lie in the examination of why a team requires a team-building intervention and the characteristics of the team. Human resources practitioners could play a more proactive role in identifying teams that could benefit from team building. Specifically, the finding that the role-clarification and goal-setting components improved performance over the other team-building components could benefit human resources practitioners and organizational managers by providing increased clarity into ways in which leaders may best direct their teams (i.e., being clear about subordinates' roles and setting goals).

The results of the current meta-analysis provide encouraging news to human resource practitioners as well as to the many users of team-building interventions. Generally speaking, the data have suggested that team building has a greater impact on some outcomes over others, and some team sizes over others. In a case where a manager suspects that his or her team may benefit from team building it would serve the manager to evaluate and identify the team's characteristics, as well as the specific problems encountered prior to intervening with team building. In other words, by communicating to practitioners the varied results that team-building interventions have on different outcomes, and considering the potential moderating influence of team size, practitioners can be better prepared when assessing what type of intervention or change effort is most appropriate for their team. The results from this study reinforce the view that not all teams will benefit from the same team-building intervention.

Study Limitations

Despite the many interesting findings from the current research integration, there are a number of limitations inherent in this study. First, one conceptual limitation of this study can be summed up by the potentially interesting and important moderator variables that we did not examine. These include team type, training duration, criterion source of measurement, and perhaps also a more exacting assessment of the level of interdependence of the teams under investigation in this integration. A proper analysis of these additional factors may represent a substantial contribution to the current knowledge base concerning team-building interventions.

There were also a number of methodological limitations inherent to this research. For one thing, only a fraction of the primary studies reported sufficient data to allow for corrections of the unreliability of criterion

measurement. Depending on the analysis in question, only 10%–30% of the effect sizes in the database had an internal consistency estimate of reliability associated with it. Related to this issue, we took the liberty of assuming that the predictors, which in this case were the team-building interventions themselves, were implemented with 100% consistency and accuracy. Unfortunately, we were not aware of a better estimate of the reliability with which these interventions are characteristically implemented. Thus, our assumption and use of perfect predictor reliability is likely an overestimate of the true reliability of these interventions.

An additional methodological concern in this research concerned the issue of having a relatively small number of correlations available for many of the subgroup analyses. Unfortunately, the strict criteria set for inclusion of primary studies in the database, combined with the relative paucity of published research in this area, somewhat limited the amount of data available to us. However, this issue has been relatively common in empirical reviews of team building (e.g., Salas et al., 1999), and the current integration still represents the most exhaustive statement currently available on this topic.

Directions for Future Research and What We Still Need to Know

This meta-analysis provides many answers that expand the literature on team-building interventions. However, there are still some questions that remain unanswered. Below we present a few recommendations for future research. Research is needed that examines the efficacy of team building for various types of teams. Especially needed are published studies of team-building interventions for management, and action teams. For example, in the strategic management literature it has been noted that top management teams are necessary partners to help set strategic direction, redesign organization architecture, and improve business processes (Allaire, 1998). It would be well served for future studies to observe the impact that these interventions may have on organizational financial outcomes when implemented in management teams. Furthermore, future research could also examine if the reason by which teams are formed has any impact on the efficacy of a team-building intervention. If a team is formed because of a directive does team building affect it differently than a team that was formed by choice?

Additionally, there is a need to further investigate the issue of team size as it relates to the efficacy of team building. It is an organizational reality

that organizations require teams of all shapes and sizes. The findings from this research, where team building appeared to work better for larger teams, were in direct contrast to those reported by Salas and colleagues (1999). Concerning the current findings, if team building does affect larger teams to a greater extent, does it do so only temporarily? Is there a greater need for follow-up development activities with large teams than small teams? Could it be that there is an interaction between team size and team type, such that team size is positively correlated with improvements in team functioning only for certain team types (e.g., production teams)? For other teams (e.g., project teams) it may be the case that team building works better with entities that are smaller in size. This issue could prove to be a fruitful area of inquiry, but is currently understudied.

Lastly, very few of the studies analyzed here investigated the effects of team building over time. Instead, most were limited to an isolated postintervention measure of performance. Thus, there is a need for researchers to investigate the results of team building over the span of the team's life. It is in this area that we feel practitioners could make the most impact. It is critical to understand if teams that participate in team building work better as their time together increases. For example, the longer teams work together the more likely it is that problems will arise. Future research could assess if the techniques that teams use to resolve issues are based on what they learned from their team-building intervention. Ideally, such research would take pre-intervention measures of performance to calculate a baseline that would then be followed by multiple measures of performance after the team-building intervention.

Concluding Remarks

Our findings are encouraging—team building improves team outcomes; that these team-development interventions are beneficial to team functioning is the good news. However, we still need to know more about team building. What are the mechanisms by which it works? What specific features are best, and how can these be best designed and implemented? In short, more in-depth evaluations are needed. If our understanding of the effectiveness and boundary conditions of team building is to be further enhanced by the type of evidence-based conclusions derived from meta-analyses—again, we need more data and more evaluations. This is our call to practitioners of team-building interventions to lead the charge in assessing the impact that team building has on teams in the field. In practice, team building comes in many

forms (i.e., games, adventures, or exercises), and is a widely used intervention; this study is an attempt at explaining if it works.

References

References marked with an asterisk (*) indicate studies included in the meta-analysis.

Allaire, P. A. (1998). Lessons in teamwork. In D. C. Hambrick, D. A. Nadler, & M. L. Tushman (Eds.), *Navigating change: How CEOs, top teams, and boards steer transformation* (pp. 113-130). Boston: Harvard Business School Press.

Applebaum, E., & Batt, R. (1994). *The new American workplace: Transforming work systems in the United States*. Ithaca, NY: IRL Press.

Aronson, E., Ellsworth, P. C., Carlsmith, J. M., & Gonzales, M. H. (1990). *Methods of research in social psychology*. Boston: McGraw-Hill.

Bantel, K. A., & Jackson, S. E. (1989). Top management innovations in banking: Does the composition of the top team make a difference? *Strategic Management Journal*, *10*, 107-124.

Beer, M. (1976). The technology of organization development. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 937-994). Chicago: Rand McNally.

Blake, R. R., & Mouton, I. S. (1964). *The managerial grid*. Houston: Gulf.

*Boss, R. W., & McConkie, M. L. (1981). The destructive impact of a positive team-building intervention. *Group and Organization Studies*, *6*, 45-56.

*Bragg, J. E., & Andrews, J. R. (1973). Participative decision-making: An experimental study in a hospital. *Journal of Applied Behavioral Science*, *9*, 727-735.

Buller, P. F. (1982). *The effects of team building-problem solving and goal setting-performance feedback on the productivity of underground metal miners: A field experiment*. Unpublished doctoral dissertation, University of Washington.

Buller, P. F. (1986). The team building-task performance relation: Some conceptual and methodological refinements. *Group and Organization Studies*, *11*, 147-168.

*Buller, P. F., & Bell, C. H., Jr. (1986). Effects of team building and goal setting on productivity: A field experiment. *Academy of Management Journal*, *29*, 305-328.

*Bushe, G. R., & Coetzer, G. (1995). Appreciative inquiry as a team-development intervention: A controlled experiment. *Journal of Applied Behavioral Science*, *31*, 13-30.

*Cohen, S. B. (1993). *Effects of attribution retraining and positive peer support training on team cohesion, performance, and participant satisfaction*. Unpublished doctoral dissertation, Hofstra University, Hempstead, NY.

DeMeuse, K. P., & Liebowitz, S. J. (1981). An empirical analysis of team-building research. *Group and Organization Studies*, *6*, 357-378.

*Dionne, S. D. (1998). *Team training and development in organizations: A multiple levels of analysis field experiment*. Unpublished doctoral dissertation, State University of New York at Binghamton, Binghamton.

Dyer, W. G. (1977). *Team building: Issues and alternatives*. Reading, MA: Addison-Wesley.

Dyer, W. G. (1987). *Team building: Issues and alternatives* (2nd ed.). Reading, MA: Addison-Wesley.

*Eden, D. (1986). Team development: Quasi-experimental confirmation among combat companies. *Group and Organization Studies*, *11*, 133-146.

Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, *6*, 161-180.

Field, A. P. (2005). Is the meta-analysis of correlation coefficients accurate when population correlations vary? *Psychological Methods*, *10*, 444-467.

*Friedlander, F. (1967). The impact of organizational training laboratories upon effectiveness of ongoing work groups. *Personnel Psychology*, *20*, 289-307.

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

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

Halebian, J., & Finkelstein, S. (1993). Top management team size, CEO dominance, and firm performance: The moderating roles of environmental turbulence and discretion. *Academy of Management Journal*, *36*, 844-863.

Hambrick, D. C., & D'Aveni, R. A. (1992). Top team deterioration as part of the downward spiral of large corporate bankruptcies. *Management Science*, *38*, 1445-1466.

*Howard, J. (1979). *Organizational team building in church-related teams*. Unpublished dissertation, California School of Professional Psychology, San Diego.

*Huang, W. W., Wei, K., Watson, R. T., & Tan, B. C. Y. (2002). Supporting virtual team-building with a GSS: An empirical investigation. *Decision Support Systems*, *34*, 359-367.

*Hughes, R. L., Rosenbach, W. E., & Clover, W. H. (1983). Team development in an intact, ongoing work group: A quasi-field experiment. *Group and Organization Studies*, *8*, 161-186.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research* (2nd ed.). Thousand Oaks, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating findings across research*. Beverly Hills, CA: Sage.

Hunter, J. E., & Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis method and findings. *Journal of Applied Psychology*, *91*, 594-612.

Indik, B. P. (1965). Operational size and member participation: Some empirical tests of alternative explanations. *Human Relations*, *18*, 339-350.

Johnson, B. T., Mullen, B., & Salas, E. (1995). Comparison of three major meta-analytic approaches. *Journal of Applied Psychology*, *80*, 94-106.

*Kimberley, J. R., & Nielsen, W. R. (1975). Organization development and change in organizational performance. *Administrative Science Quarterly*, *20*, 191-206.

Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest, 7*(3), 77-124.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, *78*, 311-328.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

*Longenecker, C. O., Scazzero, J. A., & Stansfield, T. T. (1994). Quality improvement through team goal setting, feedback, and problem solving. *International Journal of Quality and Reliability Management*, *11*(4), 45-53.

Macy, B. A., & Izumi, H. (1993). Organizational change, design and work innovation: A meta-analysis of 131 North American field experiments, 1961-1991. In W. Pasmore & R. Woodman (Eds.), *Research in organizational change and development* (pp. 235-313). Greenwich, CT: JAI.

*Miller, D. J., Jr. (1997). *The impact of low ropes course training on team development.* Unpublished doctoral dissertation, Colorado Technical University, Colorado Springs.

*Mitchell, R. (1986). Team building by disclosure of internal frames of reference. *Journal of Applied Behavioral Science*, *22*, 15-28.

*Morrison, P., & Sturges, J. (1980). Evaluation of organization development in a large state government organization. *Group and Organization Studies*, *5*, 48-64.

Mullen, B. (1987). Self-attention theory. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 125-146). New York: Springer-Verlag.

Mullen, B., Anthony, T., Salas, E., & Driskell, J. E. (1994). Group cohesiveness and quality of decision making: An integration of tests of the groupthink hypothesis. *Small Group Research*, *25*, 189-204.

Mullen, B., Brown, R. J., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, *22*, 103-122.

Mullen, B., & Copper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin*, *115*, 210-227.

Mullen, B., Salas, E., & Driskell, J. E. (1989). Salience, motivation, and artifact as contributions to the relation between participation rate and leadership. *Journal of Experimental Social Psychology*, *25*, 545-559.

Noe, R. A. (2002). *Employee training and development* (2nd ed.). Boston: McGraw-Hill.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Payne, V. (2001). *The team-building workshop: A trainer's guide*. New York: AMACOM.

Porras, J. I., & Wilkins, A. (1980). Organization development in a large system: An empirical assessment. *Journal of Applied Behavioral Science*, *16*, 506-534.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638-641.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

Rosenthal, R. (1994). Parametric measures of effect size. In. H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). New York: Russell Sage.

Rosenthal, R., & Rubin, D. (1988). Comment: Assumptions and procedures in the file drawer problem. *Statistical Science*, *3*, 120-125.

Salas, E., & Cannon-Bowers, J. A. (1997). Methods, tools, and strategies for team training. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 249-280). Washington, DC: American Psychological Association.

Salas, E., Priest, H. A., & DeRouin, R. E. (2005). Team building. In N. Stanton, H. Hendrick, S. Konz, K. Parsons, & E. Salas (Eds.), *Handbook of human factors and ergonomics methods* (pp. 48-1, 48-5). London: Taylor & Francis.

Salas, E., Rozell, D., Mullen, B., & Driskell, J. E. (1999). The effect of team building on performance: An integration. *Small Group Research*, *30*, 309-329.

Scargle, J. D. (2000). Publication bias: The "file-drawer" problem in scientific inference. *Journal of Scientific Exploration*, *48*, 91-106.

Schein, E. H. (1969). *Process consultation: Its role in organization development*. Reading, MA: Addison-Wesley.
Schein, E. H. (1999). *Process consultation revisited: Building the helping relationship*. Reading, MA: Addison-Wesley.
Schmidt, F. L., & Le, H. (2005). *Software for the Hunter-Schmidt meta-analysis methods* (Vol. 1.1). Iowa City: University of Iowa, Department of Management & Organizations.
Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs*. Boston: Houghton Mifflin.
Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420-428.
Stasser, G., Vaughan, S. I., & Stewart, D. D. (2000). Pooling unshared information: The benefits of knowing how access to information is distributed among group members. *Organizational Behavior and Human Decision Processes*, *82*, 102-116.
Sundstrom, E., DeMeuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Psychologist*, *45*, 120-133.
Sundstrom, E., McIntyre, M., Halfhill, T., & Richards, H. (2000). Work groups: From the Hawthorne studies to work teams of the 1990s and beyond. *Group Dynamics*, *4*, 44-67.
Svyantek, D. J., Goodman, S. A., Benz, L. L., & Gard, J. A. (1999). The relationship between organizational characteristics and team building success. *Journal of Business Psychology*, *14*, 265-283.
Tannenbaum, S. I., Beard, R. L., & Salas, E. (1992). Team building and its influence on team effectiveness: An examination of conceptual and empirical developments. In K. Kelley (Ed.), *Issue, theory, and research in industrial/organizational psychology* (pp. 117-153). Amsterdam: Elsevier.
*Wegenast, D. P. (1983). *Team building: An application of leadership skills to case management in child protective services*. Unpublished doctoral dissertation, City University of New York, New York City.
*Wexler, K. J. (1990). *Team building: A multidimensional analysis*. Unpublished doctoral dissertation, Hofstra University, Hempstead.
Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals. *Journal of Applied Psychology*, *75*, 315-321.
*Woodman, R. W., & Sherwood, J. J. (1980). Effects of team development intervention: A field experiment. *The Journal of Applied Behavioral Science*, *16*, 211-227.

Cameron Klein works as a survey consultant at Kenexa. Prior to joining Kenexa, Cameron was employed as an organizational and teams researcher at the Institute for Simulation & Training. His primary research interests include test development, individual and team training, and interpersonal skills. He has partnered with organizations that include the Army Research Institute, Kohl's, Kellogg's, the National Aeronautics and Space Administration, Ocean Spray, and Windstream. He is currently a doctoral candidate at the University of Central Florida.

Deborah DiazGranados is a doctoral candidate in the Industrial/Organizational Psychology program at the University of Central Florida and is a graduate research assistant at the Institute for Simulation and Training. Ms. DiazGranados received a BS in Psychology and Management from the University of Houston, and her MS in Industrial/Organizational Psychology from the University of Central Florida. Her research interests include team processes and effectiveness, training, motivation, leadership, and the multicultural issues that surround these topics.

Eduardo Salas is trustee chair and professor of psychology at the University of Central Florida where he also holds an appointment as program director for the Human Systems Integration Research Department at the Institute for Simulation and Training. His expertise includes assisting organizations in how to foster teamwork, design and implement team-training strategies, facilitate training effectiveness, manage decision making under stress, and develop performance measurement tools.

Huy Le is an assistant professor at the Department of Psychology, University of Central Florida. Dr. Le received his PhD in human resources management at the University of Iowa. His research interests include personnel selection, test development and validations, cross-cultural issues, and quantitative research methodologies (meta-analysis, Monte-Carlo simulations, structural equation modeling).

C. Shawn Burke is a research scientist at the Institute for Simulation and Training at the University of Central Florida. Her expertise includes teams, team leadership and adaptability, team training/diagnosis, and team effectiveness. Dr. Burke has over 60 publications related to the above topics and work accepted at over 100 peer-reviewed conferences. She serves as an ad hoc reviewer for several journals, including *Human Factors*, *Leadership Quarterly*, *Journal of Applied Psychology*, and *Human Resource Management*.

Rebecca Lyons is a doctoral student in the Industrial and Organizational Psychology program at the University of Central Florida. She earned a BS in Psychology in 2004 from Davidson College. Rebecca is a graduate research associate at the Institute for Simulation and Training where her research interests include individual and team training, performance measurement, multiteam systems, and simulation, with an emphasis in healthcare.

Gerald F. Goodwin is a research psychologist at the U.S. Army Research Institute (ARI) for Behavioral and Social Sciences. He received his MS and PhD in Industrial/Organizational Psychology from Pennsylvania State University. Dr. Goodwin is currently a special projects officer in ARI's Strategic Initiatives Group, with responsibility for topics related to ARI's leader and organization research programs. Dr. Goodwin's research expertise is in leadership, team effectiveness, and organizational issues in joint, interagency, and multinational contexts.