15 Principles for Measuring Teamwork: A Summary and Look Toward the Future

David P. Baker
University of Central Florida

Eduardo Salas
Naval Air Warfare Center Training Systems Division, Orlando

INTRODUCTION

Teams and, as a result, team performance measurement have received an increasing amount of attention since the 1980s (e.g., Driskell & Salas, 1991; Dyer, 1984; Foushee, 1984; Salas, Bowers, and Cannon-Bowers, 1995). There have been numerous articles on teams (see Dyer, 1984; Salas et al., 1995 for comprehensive reviews of the literature), and a number of books have been published that specifically address critical issues related to team performance (see for example, Guzzo & Salas, 1995; Swezey & Salas, 1992; Wiener, Kanki, & Helmreich, 1993). It goes without saying that teamwork has become a critical element of virtually almost all organizations.

For example, in the airline industry, team training is an integral part of pilot training (referred to as Crew Resource Management training), and in the future, team performance will be evaluated along with technical performance under the new Advanced Qualifications Program (AQP; Birnbach & Longridge, 1993). However, even with this increased emphasis on teams, we contend that there is still much to be learned about the measurement of team performance. This book documents a significant body of work that has been conducted in an attempt to understand the measurement process. It provides great insight with respect to both team performance measurement theory and outlines a number of applications and evaluations of team performance measures.

The goals of the present chapter are twofold. First, we review a series of principles for measuring team performance that we previously proposed (Baker & Salas, 1992). These principles raised a number of issues for

guiding research, so our intention is to assess the extent to which these questions have been addressed through the research presented here. Second, on the basis of our earlier principles and the chapters presented, we propose a series of updated principles for team performance measurement. We hope that these principles serve as a vehicle for constructing actual team performance measurement tools as well as guiding future research in this area.

TEAMS AND TEAM PERFORMANCE MEASUREMENT

We begin by documenting why team performance measurement research is important. First we define what a team is. We recognize that numerous definitions of a *team* have been proposed. We simply feel that it is important that readers have a common frame of reference for our discussion.

What Is a Team?

Throughout the chapters in this book, there has been a shared understanding of what defines a team and what the important characteristics of teamwork are. To summarize, teams consist of, at a minimum, two or more individuals. These individuals have specific role assignments, must perform specific tasks (i.e., taskwork), and must interact or coordinate (i.e., teamwork) to achieve a common goal or outcome (Dyer, 1984; Morgan, Glickman, Woodard, Blaiwes, & Salas, 1986; Salas, Dickinson, Converse, & Tannenbaum, 1992). In addition, teams make decisions (Orasanu & Salas, 1993), have specialized knowledge and skills (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995), and often work under conditions of high workload (Bowers, Braun, & Morgan, chapter 5, this volume; Orasanu & Salas, 1993). Finally, teams can be distinguished from small groups (Brannick & Prince, chapter 1, this volume), because teams have unique requirements for coordination and task interdependency. Teamwork characteristically involves team members adjusting to each other either sequentially or simultaneously in order to achieve team goals (Dickinson & McIntyre, chapter 2, this volume).

As documented here, in the area of team research, it is safe to conclude that there is a shared understanding of the variables that define a team. This is important from the standpoint of team performance measurement because it sets boundaries on what constitutes a team (e.g., such as the interdependency among team members), and it defines variables with respect to team inputs, team processes, and team outcomes that should be accounted for in the measurement process. Essentially, it tells us what to

measure when assessing team performance and presents a foundation on which to construct team performance measurement tools.

Why Is Measurement Important?

In chapter 1, Brannick and Prince note that capturing, defining, and measuring the interactions that are characteristic of teamwork are basic to our ability to understand the meaning of teamwork. In other words, team performance measurement research will contribute to our comprehension of the processes that define teamwork. Brannick, A. Prince, Salas, and C. Prince (1993a) outlined three reasons why team performance measurement is important. First, team theory cannot move beyond the conceptual stage without the development of psychometrically sound measurement tools. Measurement, in and of itself, will contribute to the building and validating of accurate models of teamwork. Second, without quantifiable indicators of team performance, it is hard to determine what constitutes good and poor teamwork. Such information is particularly important from the standpoint of providing performance feedback during team training (Cannon-Bowers & Salas, chapter 3, this volume). Last, measurement is vital in evaluating instructional approaches to training teams. Psychometrically sound and construct valid measures will provide an indication of the extent to which training is effective.

Although the importance of team performance measurement in understanding teamwork is well-established, there are still numerous questions as to how to assess team performance (Dyer, 1984). These questions can be organized under three global headings: what to measure, when to measure, and how to measure. With respect to what to measure, questions revolve around the appropriate unit of analysis (i.e., whether performance assessment should be made of individual team members, the team as a whole, or some combination of both), the critical skill dimensions and behaviors to assess, and the critical team knowledge structures to assess. With respect to when to measure, questions revolve around the rates at which teams mature and the appropriate time at which to capture team performance, as well as the extent to which multiple assessments need to be employed to ensure stability in the measurement process. With respect to how to measure, questions revolve around the format of the measurement device, the extent to which team performance can be objectively quantified, and the extent to which judges can accurately assess team performance.

In order to address some of these questions, we prescribed a series of six principles for guiding team performance measurement research (Baker & Salas, 1992). These principles were organized around theoretical, methodological, and psychometric issues that are important in team performance measurement and serve as a backdrop for summarizing the work presented in this book.

PRINCIPLES FOR TEAM PERFORMANCE MEASUREMENT: PROGRESS AND EMERGING PRINCIPLES

We felt that our six basic principles were important, because, as Dyer (1984) pointed out, team oriented research has been judged and will continue to be judged on the basis of the quality of the measurement techniques employed. Given that little research on team performance measurement had been conducted at that time, there was a great need for such work, and we felt that these principles in part set a foundation for structuring such investigations. In this section, we use these principles as an organizing framework to document relevant progress in the field of team performance measurement. Then, on the basis of these results, we update these principles to reflect current thinking in the field and to prescribe new avenues for future research.

Principle 1. For understanding teamwork, there is nothing more practical than a good theory (Baker & Salas, 1992).

This principle addresses the necessity for theory when conducting team performance measurement research. Specifically, we suggested that research was needed that established sound teamwork theories, and that measurement approaches should be based on this research. Understanding the knowledge, skills, and attitudes that define teamwork is critical to establishing a nomological net (Cronbach & Meehl, 1955) of the interrelationships of these variables that should be the basis for structuring measurement tools (Cannon-Bowers et al., 1995).

Progress

Early team theories attempted to establish underlying team processes and behaviors that impact team performance (Alexander & Cooperland, 1965; Boguslaw & Porter, 1962; Lanzetta & Roby, 1960; Morgan et al., 1986; Nieva, Fleishman, & Reick, 1978; Siskel & Flexman, 1962). These behaviors then were the basis for measurement tools.

More recently, theories of teamwork have evolved to include other variables (e.g., knowledge requirements, cognitive skills, etc.) in addition to team behavior (Cannon-Bowers & Salas, in press). In a comprehensive review of the team literature, Cannon-Bowers et al. (1995) defined teamwork to consist of a series of team competencies that can be distinguished from individual competencies. These researchers suggest that team competencies can be thought of as the requisite knowledge (i.e., principles and concepts underlying a team's task performance), skills (i.e., psychomotor and cognitive behaviors necessary to perform the team task correctly), and attitudes that result in effective team performance. Competencies can be generic or specific to a team or generic or specific to a task.

From the standpoint of team performance measurement, new theories about the cognitive requirements for teamwork present the most challenge. This research hypothesizes that team members develop and rely on shared knowledge structures (referred to as shared mental models) to enhance coordination and that these models are directly related to team performance (Cannon-Bowers et al., 1993; Converse, Cannon-Bowers, & Salas, 1991; Orasanu & Salas, 1993). According to Cannon-Bowers et al. (1993), shared mental models are organized bodies of knowledge that are shared across members of a team. They suggest that such models have the potential to affect teamwork at two levels. First, when communication channels are limited, shared mental models enable team members to anticipate other team member behaviors and information requirements (Converse et al., 1991). Second, shared mental models of a team task enable team members to perform team functions from a common frame of reference.

Recent work, presented in this volume, is directly related to our first principle. Of these chapters, Dickinson and McIntyre focus on team knowledge and attitudes, and Kraiger and Wenzel focus on team mental models.

With respect to team knowledge and attitudes, Dickinson and McIntyre hypothesize that not only does teamwork require the performance of critical team skills, but team members must also hold positive attitudes toward the team, receive rewards based on team goals, and possess knowledge of their own task and other team members' tasks. Effective team performance results from team members monitoring their own and other's performance, communicating with other team members, and providing feedback and back-up when needed. Therefore, team performance measures must not only focus on team behavior, but must also assess prevailing team attitudes as well as team knowledge requirements.

In terms of team cognitive skills, Kraiger and Wenzel suggest that team performance is directly related to the degree to which team members have a shared understanding of the team, the task, and the environment. These researchers document the importance of how team members acquire and represent this information (i.e., cognitive models) and how these models are shared among team members. As stated previously, the measurement of mental models and the degree to which these models are shared among team members presents one of the most challenging avenues for team performance research. These researchers advocate several strategies to pursue in this area that we discuss later.

Taken together, the research reviewed here suggests that considerable progress has been made regarding team performance theory development. Team performance theories have evolved beyond simply focusing on skills

and behaviors and now account for specific knowledge requirements on the part of team members and the degree to which this knowledge is shared among team members (Cannon-Bowers & Salas, in press).

Emerging Principles

The Whole May Truly Be Greater Than the Sum of Its Parts. To date, much of the research on teamwork has focused on specific attributes of teamwork as opposed to developing unified theories and measures of teamwork. As can be seen in the chapters in this book, some researchers focused on team behavioral skills (Komaki, chapter 11, this volume; A. Prince et al., chapter 13, this volume), some researchers focused on cognitive skills (Kraiger & Wenzel, chapter 4, this volume), and some researchers focused on team knowledge and attitudes (Dickinson & McIntyre, chapter 2, this volume). Unified theories of teamwork have been proposed (Cannon-Bowers & Salas, chapter 3, this volume; Cannon-Bowers et al., 1995; Salas et al., 1992), but unified measures of teamwork have not. Recent research suggests that such measures should attempt to assess the critical team knowledge, skill (both cognitive and behavioral), and attitude competencies (Cannon-Bowers & Salas, chapter 3, this volume). Such work is important because measuring a number of team process attributes simultaneously, in such a way that promotes understanding of the unique interactions among these variables, should provide great insight regarding the true nature and characteristics of team performance. In other words, it will only be through such research that valid models of team performance can be developed. With that in mind, we suggest the following principles:

Principle 1a: Full understanding of team performance requires behavioral, cognitive, and attitudinal-based measures.

Team Theory Has Taken Two Steps Forward, But Needs to Take One Step Back. A review of the chapters in this book and the general literature on team performance measurement suggests that a great deal of theoretical work has been completed. There have been many discussions of the variables that define teamwork and what the appropriate strategies are for measuring these attributes. However, there continues to be, in our opinion, a void with respect to actual empirical research. Theorizing about teamwork has moved ahead at a rapid pace, but there has not been enough research that validates these theories. We hope now with the introduction of the new measurement tools in this book that team researchers will begin to conduct such studies. Whereas theories of teamwork are important for guiding the development of team process measures, it is equally important that

empirical data exists to support these teamwork models. With that in mind, we suggest the following principle:

Principle 1b: The development of team performance measures must be guided, in part, by theory and, in part, by empirical research.

Principle 2: What you see may not be what you get (Baker & Salas, 1992).

Principle 2 highlights the *dynamic* nature of teamwork and the fact measurement approaches must account for this characteristic. Therefore, we argued that a single snapshot of a team's performance will likely be an insufficient measurement tool, especially if this snap shot is taken early in the team evolution and maturation process (Morgan et al., 1986; Morgan, Salas, & Glickman, 1994). This principle suggests that team performance needs to be sampled over a wide variety of occasions and conditions in order to get an accurate picture of a team.

Progress

Whereas there has been considerable theoretical progress regarding team performance, little research has focused on team development and its corresponding impact on the measurement of team performance. Early research in this area suggested that teams progressed through a linear sequence of developmental phases (Nadler & Berger, 1981). More recently, this thinking has evolved to suggest that teams develop through a variety of alternative paths (Morgan et al., 1994). For example, McGrath (1991) noted that teams may follow different paths to arrive at the same outcome. He suggested that teams engage in four modes of group activity (i.e., goal choice, means choice, policy choice, and goal attainment) with respect to three team functions (i.e., production, well-being, and member support). For each team function, teams always begin with goal choice and end with goal attainment; however, the mechanisms by which they arrive at goal attainment can vary significantly.

Morgan et al. (1994) provided a comprehensive framework for understanding team development referred to as the Team Evolution and Maturation (TEAM) model. The TEAM model describes a series of developmental stages through which newly formed teams are hypothesized to evolve. These periods of development are considered to be relatively informal and overlapping. Sharp distinctions among phases are not possible due to the dynamic situations in which teams operate. To test the TEAM model, data were collected on Navy Command Information Center (CIC)

teams going through training. The results suggested that team and task skills matured differentially and to some extent independently.

Most recently, McIntyre and Salas (1995) suggested that the most critical aspect of team development is the extent to which team members have worked together as a team. It is important to note that team experience is not synonymous with work experience or tenure, rather it involves team member experience in an intact team. These researchers point out that the fundamental aspects of teamwork (e.g., performance monitoring, feedback, closed-loop communication, backing-up behavior, team awareness, within-team interdependence) can vary within a team due to the extent of practice and training that has occurred. From the standpoint of team performance measurement, this implies that one-time measures of teamwork might result in different performance levels for different team skills, depending on the experience levels team members have with a given team. As experience increases for a given team, performance on these critical skills may change producing what appears to be unreliable results.

Research on team performance measurement needs to resolve such issues to ensure the integrity of measures that are employed. Unfortunately, little research presented here directly addressed the issue of team development and maturation, and how to account for the effects of these variables in team performance measurement. It will be recalled from our earlier work (Baker & Salas, 1992) that we called for performance measures that assessed teamwork at a variety of stages and in a variety of situations. The majority of the measures presented in this book still focused on a single snapshot of a team's behavior and do not account for team member experience. Therefore, questions can be raised regarding the extent to which teams have fully developed their teamwork capabilities in some of the empirical work presented here. In a positive light however, several chapters do note the potential effects of team development on the measurement process, and one other chapter notes how team performance measurement can contribute to team development.

With respect to the effects of team development on the measurement process, Bowers et al., in their investigation of team workload, pointed out that changes may occur in mutual team knowledge and communication behaviors as a result of team member experience. These researchers hypothesize that experience is likely to result in increased mutual knowledge among team members and a decrease in the requirements for communication. Fewer communication requirements among team members could result in decreased team workload; however, as Bowers et al. point out, empirical data do not exist to confirm this proposition. Similarly, Kraiger and Wenzel suggest that shared mental models will develop among team members over time, and therefore decrease the extent to which communication is required among members of a team.

Alternatively, with respect to the effects of team performance measurement on the team development process, Cannon-Bowers and Salas (chapter 3, this volume) delineate specific requirements for measuring team performance during training. These researchers suggest that team performance measures must support team training by providing a basis for remediation (i.e., the process by which performance deficiencies are used to design and structure subsequent instruction). These researchers go on to suggest that team performance measures in training must exist at both the team and individual levels; address processes and outcomes; describe, evaluate, and diagnose performance; and provide a basis for feedback and instructional strategy selection. In the context of this principle, this research suggests that team performance measures, appropriately designed for training, will facilitate the rate at which a team matures. In other words, effective team training, which employs accurate measurement and feedback mechanisms, may offset the extent to which team members require experience in an intact team to achieve maximum performance.

In summary, the research presented in this book provides some new insight regarding the effects of team development and maturation on the measurement process. However, while these issues are discussed at a theoretical level, no direct examination of how to account for team experience when assessing team performance was presented. We still believe that this presents a significant challenge for team performance measurement research.

Emerging Principles

Growing Up Is a Part of Life. There is significant empirical evidence to support the notion that teams evolve and mature over time; teams grow up! (McIntyre & Salas, 1995; Morgan et al., 1986; Morgan et al., 1994). Team members and the team itself pass through different stages of development with respect to both taskwork and teamwork, and these stages may occur at different rates for different team members and teams. Therefore, research needs to determine whether or not different measures of team performance are effective at different stages of team development. Furthermore, measures might be developed to predict various stages of team development. Such measures would be valuable for providing feedback to team members regarding current team performance and, as a result, facilitating the team maturation process. With that in mind, we suggest the following principles:

Principle 2a: Measures must capture the dynamic nature of teamwork.

Principle 2b: Measures and measurement tools must reflect the maturation process of a team.

Teamwork: Team Members Have To Develop a Taste For It. Empirical evidence also exists to suggest that team member experience will result in improved levels of team performance (McIntyre & Salas, 1995). In other words, team member experience with an intact team will moderate the level of team process observed and the extent to which mental models have developed among team members (Bowers et al., chapter 5, this volume; Cannon-Bowers & Salas, chapter 3, this volume; Kraiger & Wenzel, chapter 4, this volume). Under this principle then, we advocate that team process measures assess team member experience levels. Research on the this issue should determine the specific effects of team member experience on team process. Results of this research will be useful in more fully understanding team maturation and in determining the points at which it is appropriate to capture an accurate picture of the level of teamwork achieved by a team. With that in mind, we suggest the following principle:

Principle 2c: Measures must account for team member experience with a team.

Principle 3. There is no escaping observation (Baker & Salas, 1992).

This principle suggests that there is likely no practical way to escape the requirement for judges in the measurement process, especially when trying to capture team behavioral skills (e.g., team situation awareness, team leadership, etc.) as opposed to global team outcomes (e.g., the plane landed safely). We felt that measuring team behavior is perhaps the most challenging aspect of team performance measurement research. Therefore, we called for research in this area that explored the practical utility and reliability (i.e., from the standpoint of rater agreement) of various rating formats.

Progress

As noted in Principle 1, early team theories attempted to establish underlying team processes and behaviors that impact team performance, and these behaviors were the basis for measurement tools. Such tools typically require team experts to observe a team and then make judgments regarding that team's performance. These early conceptualizations of teamwork led us to conclude that observation is likely to be a critical part of the team performance measurement process, and we suggested that methodologies must be identified that result high levels of interrater agreement.

Recent theories of teamwork have evolved to include other variables (e.g., knowledge requirements, positive attitudes regarding teamwork, etc.) in addition to observable team behaviors. As noted previously, Cannon-Bowers et al. (1995) defined teamwork to consist of a series of team

competencies. These include the requisite knowledge, skills, and attitudes that are necessary for effective team performance. Such models of teamwork suggest that the requirement for observation will diminish somewhat because team knowledge, cognitive skill, and attitude competencies cannot be readily observed. Therefore, to assess these team variables, new and innovative measurement methods will have to be employed.

The chapter by Kraiger and Wenzel (chapter 4, this volume) directly tackles the problem of measuring team member mental models. Initially, these researchers attempt to delineate clearly what the construct of a shared mental model is and then outline the corresponding requirements for measuring such a phenomenon. Kraiger and Wenzel suggest that measures for shared mental models should be able to assess: (a) how team members perceive, process, or react to external stimuli; (b) how team members organize or structure task relevant knowledge; (c) common attitudes or affect for task relevant behavior; and (d) shared expectations for that behavior.

Kraiger and Wenzel propose several methods for meeting the above challenges of assessing shared mental models in teams. These include: card sorting tasks, probed protocol analysis, and structural assessment. In general, the goals of these techniques are to elicit team member knowledge structures and then to compare the similarity of these knowledge structures across team members. Shared mental models are defined to exist when the resulting knowledge structures for team members are found to be similar in nature. Kraiger and Wenzel point out that several methods may be required to capture the shared mental model construct and that research needs to be conducted that tests these measures.

In summary, new developments in team theory (e.g., team competencies, shared mental models) will likely result in additional requirements to observing team member behavior in order to understand team performance. This research suggests that tools will need to be developed that provide detailed information about the cognitive processes of team members. Such information will not be readily accessible through observations of team member behavior, but require strategies that are outlined in chapter 4. This does not mean, however, that observation will not continue to be a part of the team performance measurement process. We still feel that observation will be a significant part of assessing and providing feedback regarding team behavioral skills. We only mean to point out that our thinking has evolved on this principle to indicate that observation is only likely to be one small component of measuring a team's performance.

Emerging Principles

Seeing Should Not Be a Necessary Requirement for Believing. The general consensus across the chapters in this book, and the literature on

teams in general, suggests that teamwork is comprised of observable behaviors as well as critical team knowledge, attitudes, and cognitive skills. Therefore, measures that restrict themselves to observable team behavior are only capturing part of the picture. Team performance is far more complex and not simply represented by what team members do. To understand team performance, new measures that tap team member shared mental models and interpositional knowledge among team members must be developed and validated. This is critically important given that well developed mental models among team members may actually lead to decreases in observable team behaviors (Bowers et al., chapter 5, this volume). With that in mind, we suggest the following principles:

Principle 3a. Team performance is not simply represented by what team members do.

Principle 3b. Observation is critical for measuring and providing feedback regarding team behavioral skills.

Principle 3c. Measures that assess team member shared mental models and interpositional knowledge must be developed and validated.

Principle 4. Applications, applications, applications (Baker & Salas, 1992).

Principle 4 highlights the need for applications. We argued that team performance measures need to be developed, implemented, and evaluated for a wide variety of teams in a wide variety of settings to understand the measurement process better. We felt, and still feel, that this was one of most important principles, because data collected from these applications will be useful in guiding future measurement development and contribute to the development of teamwork theories.

Progress

In our article, we called for the development and application of team performance measures for a variety of teams in a variety of settings. At that time, most team performance measures had been developed and applied in tactical decision-making teams and aircrews; few measures had been developed and applied in other environments. We felt that this was a significant void in the field, because it was only through actual team performance measurement scale development and application that both theoretical and applied questions could be answered.

Now, as documented in the chapters in this book, great strides have been made from the standpoint of developing and applying team performance

measures. The vast majority of the authors in this volume provide detailed information on how they went about developing a team performance measurement tool to address a need in a particular context. These applications span a variety of different teams and present a number of different formats for assessing team performance. These formats include behavioral checklists for measuring team skills and behaviors, surveys for assessing team member attitudes, and advanced computer simulations for assessing specific team processes such as team decision making.

With respect to behavioral checklists, several chapters in this book present research on such measures for assessing team skills and behaviors in aircrews. Chapter 7, by Dwyer, Fowlkes, Oser, and Lane, for example, describes a method of developing team performance measures that involves determining precisely what should be done during a team task and then developing a checklist on the basis of that information. Such a scripted checklist allows observers to record specific team behaviors, and this information can be used to render a judgment regarding overall team performance. Chapter 13, by A. Prince, C. Prince, Brannick, and Salas, reviews several different measurement scales that have been designed by the Navy to capture team behavior in aircrews. These scales vary in specificity from one end of the continuum, where a scale is scripted for a specific team task, to the other end of the continuum, where a scale can be applied to a variety of tasks.

With respect to survey methods, two of the chapters in this book present information on this approach. First, chapter 8, by Hallam and Campbell, describes the development of the Campbell–Hallam Team Development Survey. The purpose of this survey was to assess team-member perceptions of how their team is doing in order to provide feedback to team members to improve team performance. The survey was designed to be applied to a wide variety of teams, and the authors report reasonably strong psychometric evidence to support the use of this instrument. Chapter 9, by Mathieu and Day, also illustrates the use of a survey to measure team performance. These researchers developed a method to assess specific teamwork variables that were characteristic of within and between-departmental functions in a nuclear power plant setting.

With respect to computer simulations, chapter 6 by Hollenbeck et al. provides extensive information on a networked software program called TIDE^2. These researchers describe how this program can be used effectively to study team decision making and present preliminary data from two empirical investigations to support the use of this simulation.

Last, several other chapters outline and/or test specific strategies for assessing specific team attributes. For example, chapter 4, by Kraiger and Wenzel, describes techniques for assessing team member mental models, and chapter 5, by Bowers et al., describes measures for assessing team

workload. These measures include properties of other measures described previously, but were designed to assess specific characteristics of teamwork.

On this basis of this research then, it is safe to conclude that significant strides have been made regarding the development and application of team performance measures. In addition, new measures are being applied in contexts outside of the military and aviation such as nuclear power plants (Toquam, MaCauly, Westra, Fujita, & Murphy, chapter 12, this volume) and theater teams (Komaki, chapter 11, this volume). Such efforts are consistent with the types of efforts we called for under this principle and should lend to significant gains in understanding the team performance measurement process.

Emerging Principles

Applications, Applications, Applications[2]. Even with all the excellent work that has been undertaken, we still feel that more team performance measures need to be developed, applied, and evaluated for different teams in different settings. In particular, we would like to re-emphasize that this principle calls for data to be collected on all new measures of team performance. To date, measures have been developed and applied, but far less data have been collected on the effectiveness of these techniques. These data are critical for understanding the psychometric properties of a measurement technique, as well as for building sound theories of team performance. With that in mind, we suggest the following principles:

Principle 4a. Team performance measures must be developed, implemented, and evaluated for a wide variety of teams in a wide variety of settings.

Principle 4b. Psychometric data must be collected on all new measures of team performance.

Applications, Applications, Applications[3]. Whereas there have been a significant number of team performance measures that have been developed to evaluate team behavioral skills, far fewer measures have been developed that assess other team competencies. Therefore, teamwork measures that assess team knowledge, attitude, and cognitive skill competencies need to be developed, applied, and evaluated. As called for under Principle 4b, data should be collected on these measures to determine their psychometric properties, as well as for constructing sound theories of team performance. With that in mind, we suggest one additional new principle:

Principle 4c. Measures that assess team knowledge, attitude, and skill competencies must be developed, applied, and evaluated.

Principle 5. Judges and measures must be reliable (Baker & Salas, 1992).

Principle 5 highlighted the fact that team performance measures must be reliable. We emphasized that reliability includes internal consistency and temporal stability of the measurement tool as well as interjudge agreement. In addition, we noted the critical role reliability plays in determining the internal validity of team performance measurement scales.

Progress

In our earlier manuscript, although we pleaded for applications of team performance measures, we also noted that these measures needed to be reliable. Furthermore, we pointed out that reliability must take place at two levels. First, when team skills and behaviors are being observed by team experts and these experts make judgments regarding team performance, the reliability of these experts must be established. Second, with respect to the measurement tools themselves, the internal consistency of these measures and their component subscales must be determined as well as the temporal stability of variables that are being assessed.

There has been some research to suggest that judges can reliably assess team skills and behaviors. Brannick and his colleagues (Brannick et al., 1993a; Brannick, Roach, & Salas, 1993b; Dwyer et al., chapter 7, this volume; A. Prince et al., chapter 13, this volume) have conducted several investigations that have targeted the psychometric properties of team performance measures. In general, these investigations have shown that raters can achieve reasonable levels of agreement. For example, Brannick et al. (1993b) found on-site raters to achieve interrater reliabilities for various team process skills to range from a low of .57 (i.e., accepting suggestions) to a high of .81 (i.e., coordination), and Brannick et al. (1993a) found interrater reliabilities to range from a low of .78 (i.e., situation awareness) to a high of .93 (i.e., adaptability). Dwyer et al. (chapter 7, this volume) have reported some the highest levels of agreement with interrater reliabilities in excess of .90.

Alternatively, evidence for the internal consistency of team performance measurement tools has been less encouraging. Analyses of both the internal consistency of team process measures and the consistency of team behavior across different yet similar simulations has produced correlations that are low in magnitude. For example, Brannick et al. (1993a) found such correlations to range from a low of .02 (i.e., adaptability) to a high of .52 (i.e., communication) when examining the consistency of team process skills across two scenarios designed to be alternative forms of each other. These results indicate potential problems with the reliability of these

measurement tools or the possibility that team performance in and of itself is unreliable. Given that the reliability data reported by Brannick and his colleagues is encouraging for the levels of interrater agreement that can be achieved and the research that has shown that teamwork evolves and matures over time, the latter proposition seems most likely.

Several chapters in the current book report additional data on the levels of interrater reliability that can be achieved with various team performance measures, though far less evidence is presented on the internal consistency of various measurement tools and the stability of various team skills. Regarding interrater reliability, Komaki (chapter 11, this volume) presents some preliminary data for her Theater Teamwork Effectiveness Measure (TTEM). Komaki suggests that any measure (i.e., team or individual) that requires judgment by an individual should demonstrate that independent raters agree on their recordings and obtain interrater reliability scores of 90% or better. However, as reported by Komaki, data collected on six theater productions by six expert raters has failed to meet this 90% agreement criterion for the TTEM. Disagreements seem to arise for particularly hard-to-define areas such as sound and light changes. For example, as Komaki notes, a director might request that the set lighting be increased and the lighting designer correspondingly increases the number of lumens. However, even the most experienced raters cannot discern such a change the next night when the team's performance is observed. Komaki reports that revisions to the TTEM continue to achieve this 90% agreement criterion.

The chapter by Toquam et al. also presents interrater reliabiiity for their investigation of nuclear power plant teams. Here, seven expert raters provided team performance ratings on three dimensions; team unity, team spirit, and team performance. For each dimension, raters were provided with guidelines for the types of behaviors to use to guide their ratings. Analysis of the data showed, across different pairs of raters, that interrater reliability levels ranged from .62 to .97, with the mean reliability across all pairs of raters being .80.

Regarding internal consistency and stability team performance measures, Hallam and Campbell (chapter 8, this volume) present fairly comprehensive data for their Team Development Survey™ (TDS™). The TDS™ was designed to measure perceptions of team members and feed back this information to team members in a form they can use for identifying strengths and weaknesses. In other words, this survey was designed to assess how a team is doing and ways a team can improve its performance. Data on the reliability of this measure, collected from a wide variety of teams, has been encouraging. Hallam and Campbell report the scale scores of the TDS™ to have a median alpha of .69 and a median test–retest reliability of .80.

In summary, psychometric data are beginning to become available for measures that target team behavioral skills, but few data are available on measures that target team knowledge, attitude, and cognitive skill competencies. It appears that observers can, in fact, be trained to achieve high levels of agreement when evaluating team behaviors, but questions still remain regarding the internal consistency of team performance measures and the extent to which team performance in itself is reliable. Answers to these questions are vital precursors to understanding the requirements for valid team performance measurement. Generalizability analysis is a potentially powerful technique for pursuing such research (Mathieu & Day, chapter 9, this volume).

Emerging Principles

Judges, Measures, and Team Performance. Here, we wish to modify our earlier principle so that it applies to team performance measures in a wider variety of contexts, especially when observation is not a requirement of the measurement process. We still feel in all team performance measurement research that information on the reliability of such measures should be collected. We simply wish to point out that reliability studies should be designed to reflect the specific characteristics of a measurement tool. Moreover, we advocate the researchers in this area look to techniques such as generalizability analysis (for an example of an application of this technique refer to Mathieu & Day, this volume) to understand the unique contributions of different variance components in the measurement process. With that in mind, we suggest the following principles:

Principle 5a. Reliability studies must reflect characteristics of the measurement tool.

Principle 5b. Team performance expert observers must demonstrate high levels of agreement (around 90%).

Principle 5c. Team performance measures must demonstrate internal consistency.

Past Performance May Not Be the Best Indicator of Future Performance. In addition to the reliability of team performance measurement tools, the reliability of team performance itself must be established. This may be difficult, because the team maturation literature suggests that team performance is not consistent over time and these changes are likely to vary across teams. Research in this area might initially focus on trying to distinguish those team skills that possess temporal stability from those that do not. With that in mind, we suggest the additional new principle:

Principle 5d. Measures must establish the reliability of team performance.

Principle 6. Validation for practice and theory (Baker & Salas, 1992).

This principle highlighted the fact that team performance measures must be valid. As noted earlier, validation is important for both applied and theoretical work in this area. From an applied standpoint, the development of valid team performance measures provides accurate information for the process of evaluating teamwork skills and conducting team training. From a theoretical standpoint, valid teamwork measures provide additional information on the knowledge, attitude, and skill competencies that underlie team process.

Progress

Few researchers have conducted extensive investigations into the validity of teamwork measures. Although the chapters in this book reflect extensive research from the standpoint of developing and applying actual measurement tools, little validation evidence is presented. However, we wish to temper this criticism somewhat by pointing out that we do recognize that team performance measures must be developed before actual validation research can be conducted. In addition, a number of the authors in this book point out the need for validating the measures they have reviewed or proposed and strategies for doing such research (see, e.g., for example Cannon-Bowers & Salas, chapter 3, this volume).

As with the research on reliability, Brannick and his colleagues have conducted the most significant research on the validity of team performance measures (Brannick et al., 1993a; Brannick et al., 1993b; A. Prince et al., chapter 13, this volume). To meet the psychometric requirements for valid team performance measurement, Brannick et al. (1993a) suggested that at a minimum, team process measures should: (a) be reliable in the sense that different judges should be interchangeable for one another, (b) be sensitive to differences in teams existing prior to task performance, and (c) be useful in predicting important team outcomes.

Brannick and his colleagues have attempted to address these issues by exploring both the reliability and construct validity of several team process measures. In general, these studies have demonstrated reasonable levels of interjudge agreement (refer to the research presented under Principle 5) and mixed results regarding construct validity. Investigations of construct validity have characteristically examined the extent to which team performance for a number of team process skills are consistent across alternate yet similar scenarios. Results of these studies have shown evidence of convergent and discriminant validity; however, method variance typically overshadows these results.

Two chapters in this volume present preliminary data for the validity of various team performance measures. First, Hallam and Campbell (chapter 8, this volume) assessed the extent to which the TDS was related to three indications of the team's performance: an average independent observer performance rating, an average team-leader performance rating, and an average team member performance rating. Results of these analyses showed that the highest correlations between the three measures of performance and the TDS scores were for the Skills scale, Commitment scale, Innovation scale, and the Leadership scale. Correlations among these performance criteria ranged from a low of .36 to a high of .66 across these scales.

Second, Toquam et al. (chapter 12, this volume) examined the extent to which characteristics of the task, characteristics of crew members, and characteristics of crew processes influence crew performance variability in nuclear power plant operations. Team performance data were collected on operator teams from seven Japanese nuclear power utilities. Regarding the crew process data, analysis focused on communication within a team and its relationship to team performance variability on a simulated team task. The results showed that crews that were not proficient in sharing technical information throughout the team demonstrated higher levels of performance variability. In other words, teams with poor communication skills performed more poorly on a team task than crews with stronger team communication skills.

In summary, the research presented in this book outlines a significant body of work that develops and applies new team performance measurement tools, and these researchers are now beginning to examine both the construct- and criterion-related validity of these measures.

Emerging Principles

Content and Construct Validation. Content and construct validation strategies will be particularly important for building team theories. Content validation strategies ensure that appropriate information is being sampled regarding specific team constructs (i.e., knowledge, skills, and attitudes), and construct validation strategies will provide empirical evidence to support the existence of those constructs. Collectively, this research should provide important information regarding team cognitive and behavioral skills, as well as the interrelationships of these skills. In other words, these studies will be vital for truly identifying the nomological net of variables that define team performance and determining how a team works most effectively. With that in mind, we suggest the following principles:

Principle 6a. The content and construct validity of team performance measures must be determined.

Principle 6b. Valid team performance measure must contribute to the development of valid team performance theories.

Criterion-Related Validation. Criterion-related validation strategies are particularly important for understanding the extent to which team process is related to team outcomes. To date, most measurement research has focused on team constructs and, to a lesser extent, the relationship between team process and performance. Therefore, we suggest that criterion-related validation studies be undertaken that examine the extent to which team cognitive and behavioral skills are related to independent team performance criteria. Such research will make a valuable contribution in the area of team training, where evaluation and feedback are viewed as critical. However, this work should not exclude examining the underlying strategies by which teams utilize their coordination skills, because different coordination strategies may result in the same level of team performance, yet be variously efficient. For example, in certain cases stronger team members may compensate for weaker team members. In such cases, the level of performance achieved might be quite high but the efficiency by which such a team reaches this level of performance might be very low. With that in mind, we suggest the following principles:

Principle 6c. The criterion-related validity of team performance measures must be determined.

Principle 6d. Team performance measures must predict team outcomes.

Face Validation. Last, team performance measures should possess face validity; they should look like they assess team performance. This issue is particularly important from the standpoint of team training and the degree to which trainees are receptive to feedback. Therefore, team performance measures should appear to be valid by targeting specific team skills that are derived from a thorough analysis of the team in question. With that in mind, we suggest the following principle:

Principle 6e. Team performance measures must look like they assess team performance.

LOOKING TOWARD THE FUTURE

In concluding, we suggest that the future of team performance measurement looks quite bright based on the significant body of research presented

throughout the chapters in this book. Since we advocated our initial list of principles (Baker & Salas, 1992), team theory has advanced at a rapid pace; a number of new team performance measures have been developed and applied in a variety of contexts; and psychometric research has started to explore the properties of team performance measures, and continues to grow. This recent research was the foundation for an additional 20 principles that we presented in this chapter to guide future team performance measurement research. These emerging principles, as well as our original six principles, are presented in Table 15.1. With these new principles as the guiding framework, we see three major trends for the future of team performance measurement research over the next several years.

Unified Theories of Teamwork

First, we envision that future research on teams and team performance will see the development of unified theories of teamwork. Team theories will no longer be simply attitude-based, knowledge-based, or skill-based, but will account for all of these variables in a single model. In fact, some of this research has already begun. For example, Cannon-Bowers and her colleagues (Cannon-Bowers et al., 1995; Cannon-Bowers & Salas, chapter 3, this volume) defined *teamwork* to be a function of the environment in which the team operates and a set of critical team member attitude, knowledge, and skill competencies.

In this arena, we also envision that the future will bring more intricate theories of team member cognitive skills and the mechanisms by which team members develop shared mental models. Research on the cognitive aspects of teamwork and the measurement of these variables is really in the earliest phases and is likely to grow at a significant rate in the future (Kraiger & Wenzel, chapter 4, this volume).

New Measurement Development

Second, we envision that the future of team performance measurement will see continued growth in the area of developing team performance measurement tools. As we already pointed out under our original principles (i.e., Principle 4: Applications, Applications, Applications), this is one of the most critical areas, if team performance measurement research is going to continue to develop as a science. To date, the progress in this area has been outstanding and it appears that the future will see continued advancement in this area both with respect to measurement design and application. As noted throughout this chapter, this work is particularly important, because it is only through the application of team performance measures that

TABLE 15.1
Principles for Measuring Teamwork Skills

Original Principles	*Emerging Principles*
1. For understanding teamwork, there is nothing more practical than a good theory (Baker & Salas, 1992).	1a: Full understanding of team performance requires behavioral, cognitive, and attitudinal-based measures. 1b: The development of team performance measures must be guided, in part, by theory and, in part, by empirical research.
2. What you see may not be what you get (Baker & Salas, 1992).	2a: Measures must capture the dynamic nature of teamwork. 2b: Measures and measurement tools must reflect the maturation process of a team. 2c: Measures must account for team member experience with a team.
3. There is no escaping observation (Baker & Salas, 1992).	3a. Team performance is not simply represented by what team members do. 3b. Observation is critical for measuring and providing feedback regarding team behavioral skills. 3c. Measures that assess team member shared mental models and interpositional knowledge must be developed and validated.
4. Applications, applications, applications (Baker & Salas, 1992).	4a. Team performance measures must be developed, implemented, and evaluated for a wide variety of teams in a wide variety of settings. 4b. Psychometric data must be collected on all new measures of team performance. 4c. Measures that assess team knowledge, attitude, and skill competencies must be developed, applied, and evaluated.
5. Judges and measures must be reliable (Baker & Salas, 1992).	5a. Reliability studies must reflect characteristics of the measurement tool. 5b. Team performance expert observers must demonstrate high levels of agreement (around 90%). 5c. Team performance measures must demonstrate internal consistency. 5d. Measures must establish the reliability of team performance.
6. Validation for practice and theory (Baker & Salas, 1992).	6a. The content and construct validity of team performance measures must be determined. 6b. Valid team performance measure must contribute to the development of valid team performance theories. 6c. The criterion-related validity of team performance measures must be determined. 6d. Team performance measures must predict team outcomes. 6e. Team performance measures must look like they assess team performance.

validated theories of teamwork can develop. The future, then, is likely to bring new measures that contribute to building unified theories of teamwork and advancing our understanding of team member cognitive skills and shared mental models.

More Validation Research

Last, we envision that the future of team performance measurement research will bring more detailed investigations of the psychometric properties of team performance measures. This research will be a direct result of the development and application of new team performance measurement tools for a variety of teams in a variety of contexts. As more and more data become available through repeated applications, researchers will begin to examine the extent to which these measurement devices are reliable and valid. It is our hope that reliability studies will seek to partial out the variance that can be attributed to different components of the measurement process (e.g., observers, rating formats, etc.) and that validity studies will seek to establish the construct and criterion-related validities of team performance measures. The majority of chapters in this book note the importance of conducting psychometric studies and outline future plans for such investigations. Therefore, we believe that the next several years will see a significant growth in research in this area.

SUMMARY

As we concluded in our earlier manuscript, some strides have been made, but there is still much to be learned about the measurement of team performance. We hope that these revised principles present an up-to-date framework for conducting such research.

Furthermore, we hope that researchers capitalize on the massive amounts of diverse information on team performance measurement that has been presented throughout this book. We believe the future of team performance measurement to be bright, but there are still many lessons to be learned here and many questions have been raised that need to be answered.

REFERENCES

Alexander, L. T., & Cooperland, A. S. (1965). *System training research in team behavior* (TM-2581; DTIC No. AD620606). Santa Monica, CA: System Development Corporation.

Baker, D. P., & Salas, E. (1992). Principles for measuring teamwork skills. *Human Factors, 34*, 469–475.

Birnbach, R. A., & Longridge, T. M. (1993). The regulatory perspective. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 263–281). New York: Academic Press.

Boguslaw, R., & Porter, E. H. (1962). Team functions and training. In R. M. Gagne (Ed.), *Human performance and productivity* (pp. 387–416). New York: Holt, Rinehart, & Winston.

Brannick, M. T., Prince, A., Salas, E, & Prince, C. (1993a, April). *Impact of raters and events on team performance measurement*. Paper presented at the eighth annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Brannick, M. T., Roach, R. M., & Salas, E. (1993b). Understanding team performance: A multimethod study. *Human Performance, 6*(4), 287–308.

Cannon-Bowers, J. A., & Salas, E. (in press). Teamwork competencies: The interaction of team member knowledge, skills, and attitudes. In H. F. O'Neil (Ed.), *Workforce readiness: Competence and assessment* Mahwah, NJ: Lawrence Erlbaum Associates.

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision making. In N. J. Castellan, Jr. (Ed.), *Current issues in individual and group decision making* (pp. 221–246), Hillsdale, NJ: Lawrence Erlbaum Associates.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. In R. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 333–380). San Francisco: Jossey-Bass.

Converse, S. A., Cannon-Bowers, J. A., & Salas, E. (1991). Team member shared mental models: A theory and some methodological issues. *Proceedings of the Human Factors Society 35th Annual Meeting*, 1417–1421.

Cronbach, L. J., & Meehl, P. H. (1955). Construct validity of psychological tests. *Psychological Bulletin, 89*, 281–302.

Driskell, J. E., & Salas, E. (1991). Group decision making under stress. *Journal of Applied Psychology, 76*, 473–478.

Dyer, J. L. (1984). Team research and training: A state-of-the-art review. In F. A. Muckler (Ed.), *Human factors review* (pp. 285–323). Santa Monica, CA: Human Factors and Ergonomics Society.

Foushee, H. C. (1984). Dyads and tryads at 35,000 feet: Factors affecting group process and aircrew performance. *American Psychologist, 39*(8), 885–893.

Guzzo, R., & Salas, E. (1995). *Team effectiveness and decision making in organizations*. San Francisco: Jossey-Bass.

Lanzetta, J. T., & Roby, T. B. (1960). The relationship between certain group process variables and group problem-solving efficiency. *Journal of Social Psychology, 52*, 135–148.

McGrath, J. E. (1991). Time, interaction, and performance: A theory of groups. *Small Group Research, 21*, 147–174.

McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Lessons from complex environments. In R. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 149–203). San Francisco: Jossey-Bass.

Morgan, B. B., Glickman, A. S., Woodward, E. A., Blaiwes, A. S., & Salas, E., (1986). *Measurement of team behaviors in a Navy environment* (Tech. Rep. No. NTSC TR-86-014). Orlando, FL: Naval Training Systems Center.

Morgan, B. B., Salas, E., & Glickman, A. S. (1994). An analysis of team evolution and maturation. *The Journal of General Psychology, 120*, 277–291.

Nadler, L., & Berger, L. (1981). *A classification system for Navy teams* (final report). (DTIC No. AD A-105 400). Alexandria, VA: Mellonics Systems Development Division.

Nieva, V. F., Fleishman, E. A., & Reick, A. (1978). *Team dimensions: Their identity, their measurement, and relationships* (Final Tech. Rep., Contract DAH19-78-C-0001). Washington, DC: Advanced Resources Research Center.

Orasanu, J., & Salas, E. (1993). Team decision making in complex environments. In G. Klein, J. Orasanu, & R. Calderwood (Eds.), *Decision making in action: Models and methods* (pp. 327–345). Norwood, NJ: Ablex.

Salas, E., Bowers, C. A., & Cannon-Bowers, J. A. (1995). Military team research: 10 years of progress. *Military Psychology, 7*(2), 55–75.

Salas, E., Dickinson, T. L., Converse, S., & Tannenbaum, S. I., (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams their training and performance* (pp. 3–29). Norwood, NJ: Ablex.

Siskel, K., & Flexman, R. (1962). *Study of the effectiveness of a flight simulator for training complex aircrew skills*. Unpublished data, Bell Aeronautics Company.

Swezey, R. W., & Salas, E. (1992). *Teams their training and performance*. Norwood, NJ: Ablex.

Wiener, E. L., Kanki, B. G., & Helmreich, R. L. (1993). *Cockpit resource management*. New York: Academic Press.

SERIES IN APPLIED PSYCHOLOGY

n A. Fleishman, George Mason University
Series Editor

work and the Bottom Line: Groups Make a Difference
Rosen

rns of Life History: The Ecology of Human Individuality
ael D. Mumford, Garnett Stokes, and William A. Owens

Motivation
E. Kleinbeck, Hans-Henning Quast, Henk Thierry, and Hartmut Häcker

iology in Organizations: Integrating Science and Practice
n R. Murphy and Frank E. Saal

an Error: Cause, Prediction, and Rèduction
W. Senders and Neville P. Moray

mporary Career Development Issues
rt F. Morrison and Jerome Adams

e in the Workplace: Approaching Fairness in Human Resource
gement
ll Cropanzano

nnel Selection and Assessment: Individual and Organization
ectives
Schuker, James L. Farr, and Mike Smith

nizational Behavior: The State of the Science
l Greenberg

e Psychology Into the 21st Century
n I. Kurke and Ellen M. Scrivner

hmark Tasks for Job Analysis: A Guide for Functional Job
sis (FJA) Scales
y A. Fine and Maury Getkate

and Human Performance
s E. Driskell and Eduardo Salas

oving Training Effectiveness in Work Organizations
vin Ford, Steve W. J. Kozlowski, Kurt Kraiger, Eduardo Salas,
Mark S. Teachout

Feedback: Giving, Seeking, and Using Feedback for Performance Improvement
al London

Performance Assessment and Measurement: Theory, Methods,
pplications
ael T. Brannick, Eduardo Salas, and Carolyn Prince

TEAM PERFORMANCE ASSESSMENT AND MEASUREMENT

Theory, Methods, and Applications

Edited by

Michael T. Brannick
University of South Florida
Eduardo Salas
Carolyn Prince
Naval Air Warfare Center, Training Systems Division

LEA LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1997 Mahwah, New Jersey London