

**Running head: PERFORMANCE MEASUREMENT IN SIMULATION**

**Performance Measurement in Simulation-Based Training: A Review and Best  
Practices**

Eduardo Salas  
Michael A. Rosen

Department of Psychology, and  
Institute for Simulation and Training  
University of Central Florida

Prepared for:  
Janet D. Held  
Navy Personnel Research, Studies, & Technology (NPRST/PERS-13)  
Bureau of Naval Personnel  
5720 Integrity Drive  
Millington, TN 38055-1000

31-07-2007

Address correspondence to:  
Eduardo Salas, Ph. D.  
Professor & Trustee Chair  
Department of Psychology, and  
Institute for Simulation & Training  
University of Central Florida  
3100 Technology Parkway  
Orlando, Fl 32826  
(407) 882-1325  
(407) 882-1550 fx  
(407) 970-1375 cell  
[esalas@ist.ucf.edu](mailto:esalas@ist.ucf.edu)

**Abstract**

Simulation-based training (SBT) is a methodology for providing systematic and structured learning experiences. The effectiveness of this methodology is dependent upon the quality of performance measurement practices in place. Performance measurement during SBT must be diagnosed; that is, the causes of effective and ineffective performance must be determined. This diagnostic measurement drives the systematic decisions concerning corrective feedback and remediation. The purpose of this paper is to provide a state of the science review of human performance measurement systems in SBT. To this end three specific goals are addressed. First, a review of the theoretical foundations being used to drive performance measurement systems in SBT is provided. Second, an overview of the methodologies and approaches to measurement in SBT is provided. Third, a set of best practices for designing performance measurement systems for use in SBT are provided. These best practices are based on the scientific and practice-based literatures.

### **Why is Performance Measurement Important?**

In domains such as the military, medicine, business and aviation, the present day workplace is characterized by high degrees of complexity and competitiveness. In this environment, maximizing human performance is essential for safety, effectiveness, and organizational survival. Human performance can be resilient, adaptive, and flexible in ambiguous and information intensive contexts, but it can also be plagued with error and inefficiencies. The measurement of performance is key to managing, elevating, and sustaining levels of human performance. Simply put, if performance is not defined and measured, it can not be changed or improved in a systematic way. There are many goals, purposes, and reasons to measure performance in simulation and training environments. These include, but are not limited to supporting personnel selection, diagnosis, feedback in the training process, assessment of training outcomes, assessment of interventions, and development. Generally, performance measurement can be viewed as an investment in time and monetary expense that pays dividends in information upon which decisions can be made and action can be taken (Brannick & Prince, 1997). Whether dealing with team or individual performance, the ability to accurately measure (and diagnose) is a necessity to effectively meet any of the above-mentioned goals of performance measurement. Consequently, an unformed approach to performance measurement not only squanders the investment of time and other resources required to implement a performance measurement system, but may incur additional cost in terms of poor decisions being made and improper actions being taken.

Therefore, this paper sets out to answer a series of questions regarding the state of the science in human performance measurement systems. First, what are the theories driving performance measurement research and practice? What are the theoretical foundations behind the development of performance metrics and measurement systems? Second, what methodologies are being used? An examination of the existing tools provides insight into the breadth of measurement techniques available for use today. Third, what are the ‘best practices’ in current performance measurement systems? Answering these questions requires examining the requirements of performance measurement systems and assessing the degree to which a performance measurement, technique or measurement system fulfills these requirements.

### **What Theories are Driving Performance Measurement Research?**

Effective performance measurement cannot be divorced from theories of the underlying constructs being measured. However, the assertion that the purpose of the measurement system guides the selection of measures and methods of measurement (Cannon-Bowers et al., 1989) implies that a broad repertoire of theories will be called upon in the development of performance measurement systems. The selection of a theoretical grounding for performance measurement changes with context of use and purpose of measurement. This section reviews several of the major theories used in the development of team and individual performance measurement systems.

#### *Individual Performance*

Performance measurement on the individual level has been driven by numerous theories. This section provides an overview of some of the more influential perspectives

including human information processing, stress and mental workload, human expertise, and general theories of human performance.

*Human Information Processing (HIP)*

Much of the current understanding of human performance on the individual level is rooted in theories of human information processing, especially for the large and increasing number of cognitive tasks that dominate current work settings. There are numerous instances of HIP models of human performance (e.g., Card, Moran, & Newell, 1983; Wickens, 1984) stemming from early discrete stage models (e.g. Sternberg, 1969; Broadbent, 1972; Posner, 1978). The essence of HIP models of human performance is that information in different forms flows from the environment or system (comprising people and technology) through the individual. The individual makes different types of transformations on this information; these transformations are the sources of effective and ineffective performance (Wickens & Carswell, 1997). Two representative HIP models of human performance, the Wickens model and the Recognition Primed Decision making model, are discussed in more detail.

*The Wickens Model.* The Wickens model (Wickens, 1984; 1987; Wickens & Hollands, 2000) is a synthesis of resource theory and human information processing theory. Resource theory posits that in order to perform cognitive tasks, humans draw on pools of resources and that these resources are limited. The Wickens model differentiates the available resources so that demands (i.e., mental workload) may be distributed across them in a manner that maximizes performance efficiency. This model outlines different pools of mental resources that can be called on separately in order to process different types of tasks in parallel.

The Wickens model uses three separate dimensions (i.e. the stages of information processing, the modality of input, and the coding of information processing) to define the different resources available in the human information processing system. The stages of information processing dimension consists of perception, working memory (central processing), and response processes; that is, the model proposes that these types of information processing represent different pools of cognitive resources available. Therefore, at any one time a task requiring response selection resources should not interfere with a task that requires encoding resources. However, larger scale tasks will usually require all types of processing. Hence, these three stages are further divided by manner of coding. The coding of information processing consists of spatial and verbal; these codes apply across all three levels of the information processing stages (i.e. verbal coding in the central processing stage is a different resource pool than the spatial coding in the central processing stage.) As a design heuristic for determining when tasks will or will not interfere, tasks comprised of “moving, positioning, or orienting objects in space or performing other analog transformations” (Wickens, 1987, p. 97) are considered spatial and tasks mediated by “words, language, or logical operations” (Wickens, 1987, p. 97) are verbal. The modality of input dimension consists of visual and auditory modalities and output dimensions are vocal and manual. The Wickens Model provides an elegant synthesis of a number of psychological theories and knowledge in a way that transcends (or facilitates the transcendence) of the gulf between theory and practice, knowledge and its application in design and evaluation.

*The Recognition Primed Decision (RPD) Model.* The RPD Model (Klein, 1993) represents a different type of HIP model, one developed from an ecological perspective

of HIP in natural settings outside the laboratory (Wickens & Carswell, 1997). This type of model fits in a class of macro-cognitive models (Klein, Ross, Moon, Klein, Hoffman, & Hollnagel, 2003; Cacciabue & Hollnagel, 1995) which are characterized by a focus on cognitive work as it naturally occurs in situations involving complex decisions, time pressure, high risk, high stakes, ill-defined goals, and conditions where all key variables are not under control of the individual. The RPD model describes an expert's decision-making process as being comprised of two general sub-processes: (1) cue and pattern recognition, and (2) mental simulation of courses of action. First, a decision-maker senses cues in the environment and attempts to match the pattern of cues in the decision environment with past experiences. If the match for the present problem is found, the decision-maker retrieves the following: (1) associated expectancies about the problem, (2) possible courses of action that are likely to be effective, (3) a set of cues that are most relevant to the problem, and (4) goals for the situation. If a reasonable match is not immediately found, the expert decision-maker engages in a more exhaustive diagnosis of the situation which involves gathering more informational cues about the problem.

Once the situation has been assessed and relevant information retrieved from memories of similar past experiences, the expert decision-maker mentally simulates a course of action to determine if the solution will work given the contextual constraints which characterize the current situation. The decision-maker's mental simulation results in either the adoption of the original course of action, modification of particular aspects of the retrieved course of action so that it is more effective in the present situation, or rejection of the course of action in lieu of retrieval and mental simulation of an alternate course of action. The RPD model makes explicit the crucial role of situation assessment

and situation awareness in expert decision-making processes in natural real world settings. Endsley (1997) proposes that the three levels of situation awareness (i.e. perception of elements in the environment, comprehension of the current situation, and projection of future status) are the inputs to the decision-making process. That is, processes culminating in an understanding of the current and likely future states of the environment feed directly into the decision-making process. This process is essential to the subsequent cue recognition and retrieval of expectancies, goals, relevant cue sets, and possible courses of action stages of the RPD model.

### *Stress and Mental Workload*

In an attempt to remedy the incongruence of strict HIP models with the reality of human performance in a real world context, human factors researchers have addressed energetic aspects of HIP (e.g. Driskell & Salas, 1996; Hancock & Meshkati, 1988; Hockey, 1984; 1986). That is, the impact of factors such as stress, workload and fatigue on the human IP system are now viewed as essential additions to the understanding of human performance in dynamic environments. Research into energetics and performance has a long history, the beginning of which will be taken to be the unitary concept of arousal. A famous application of this concept to performance is the Yerkes Dodson Law (Yerkes & Dodson, 1908). Though their work was based on experimentation with a small number of mice, their ‘inverted U’ model of stress effects on performance neatly summarized the general arousal perspective and therefore achieved longevity as a general explanation. Specifically, starting at low levels of arousal (stress), increasing arousal will increase performance up until the optimal level. After this point, increasing arousal has deleterious effects on performance by impairing functions like attention and memory

(Wickens & Hollands, 2000). However, not all of the empirical evidence supports a one-dimensional construct of arousal. Notably, different stressors impact performance on a single task differently, and a single stressor differentially effects performance on different types of tasks.

Therefore, recent conceptualizations of stress effects on performance have fragmented the unitary concept of arousal. The effect of stress on a system is a multilevel phenomenon in several senses. First, differential effects of stress type (source) suggest that “the nature of stress effects can only be properly understood by considering the particular patterns of bodily and cognitive changes associated with different stressors as indicators of specific stress states” (Hockey & Hamilton, 1983, p. 331). This approach has led to a broad-band analysis (research investigating the effect of one stressor over many task types, Hockey & Hamilton, 1983) of stress using several indicator variables (e.g. alertness, selectivity, speed and accuracy, and STM capacity). These variables represent cognitive functioning and are of one of two types: strategic and structural. Strategic variables are those that are “particular to the task situation and its demand” (Hockey & Hamilton, 1983, p. 347) and structural variables “reflect fundamental changes in the operating parameters of the system” (Hockey & Hamilton, 1983, p. 347). By examining the effects that different types of stress have on these outcome variables, it is possible to sketch a stress state map, wherein the effects of different stress types can be organized. These stress state maps are descriptive of the performance effects of stress without imposing a strict theory of causation.

A more integrative and process focused approach to the effects of stress on attentional resources is provided by the Hancock and Warm (1989) Model of Stress and

Sustained Attention. This model seeks to integrate into one process three disparate views of stress: input (stress as physical characteristics of the environment), adaptation (stress as compensatory processes of the body to accommodate environmental change), and output (stress related to goal-directed behavior). Briefly, the model of stress and performance (sustained attention specifically) is comprised of three states, one which represents the successful achievement of dynamic stability, the second which represents dynamic instability that increases to a state of failure, and the third which is a state of transition between the first and second states. Essentially, the model states that as input stress increases, the system adapts through psychological and physiological processes until the stress levels increase to the point where adaptive capacity is exceeded (either at the point of extreme overload—hyperstress—or extreme underload—hypostress).

### *Expertise*

The study of human expertise—of reliably superior performance—has a long history in the scientific community (Ericsson, 2006). Traditionally, this research has taken place in task domains such as chess, sports, and music. However, expertise research increasingly makes contributions to the understanding of performance in complex tasks within organizations in two ways. First, theoretical work concerning the nature of expertise developed in tightly defined tasks has contributed to an understanding of performance in less defined tasks (e.g., those usually trained with SBT; Salas & Rosen, in press). Second, more cognitively complex ‘real world’ tasks are more frequently becoming the subject of expertise research (e.g., Sonnentag, Niessen, & Volmer, 2006). For example, the RPD model described earlier is rooted in the expertise literature and was first developed to describe the performance of firefighters. Expertise is best

understood as a set of psychological and physiological adaptations to the constraints of the performance task (Ericson & Lehman, 1996). That is, there is no general theory of expertise because the nature of expertise is different depending on the domain of expertise (Ericsson & Smith, 1991). However, the extensive empirical literature on expert performance across different domains has been synthesized to reveal a set of characteristics that contribute to expertise to varying degrees depending on the nature of the task (see Glaser, 1987; Glaser & Chi, 1988; Bedard & Chi, 1992; Chi, 2006; Feltovich, Prietula, & Ericsson, 2006; Hoffman et al., 1997). These characteristics, or mechanisms of expertise, are illustrated in Figure 1 and should be considered as a ‘prototype’ of expertise (Sternberg, 1997) where each of these mechanisms may contribute to expert performance in a given domain depending on the task constraints. In the simplest of cases, expert performance can be viewed as a function of extended experience within a domain and intentional practice activities that yield a larger more organized knowledge base (Bordage & Zachs, 1984; Chi, Feltovich, & Glaser, 1981) and memory skill (Ericsson & Kintsch, 1995), which in turn facilitates pattern recognition and problem representation abilities (Simon & Chase, 1973; Glaser & Chi, 1988; Zeitz, 1997), as well as self-monitoring skills and automaticity (Moors & De Houwer, 2006; Lesgold, Rubinson, Feltovich, Glaser, Klopfer, & Wang, 1988).

-----  
Insert Figure 1 about here.  
-----

*Theory of Performance and Performance Determinants*

Campbell and colleagues have advanced a theory positing that performance is a multidimensional construct comprising distinct components (e.g. job-specific task proficiency, written and oral communication, facilitating peer and team performance) and that that relevant variance in these distinct performance components can be accounted for by three factors, or determinants of performance (Campbell, McCloy, Oppler, & Sager, 1993). According to their model, performance is a function of (1) declarative knowledge, (2) procedural knowledge and skill, and (3) motivation. Declarative knowledge is defined as the facts, principles, goals, and self-knowledge necessary to complete a given task and more broadly can be considered an understanding of the task requirements. Procedural knowledge and skill involves the cognitive, psychomotor, physical, self-management, and interpersonal skills relevant to a task; these skills are attained when knowing what to do (declarative knowledge) is combined with knowing how it is done. Campbell and colleagues define motivation as a result of three choice behaviors including the choices concerning the expenditure of effort, the level of effort, and the persistence of effort. This theory has received direct empirical support, although the model tested assumed that the three performance determinants functioned in an additive manner, when a multiplicative relationship is more likely (McCloy, Campbell, & Cudeck, 1994). For example, an additive model suggests that acceptable levels of performance can be achieved with high levels of two determinants (e.g. declarative knowledge, and procedural knowledge and skill) and none of the third determinant (e.g. motivation). In a multiplicative model there must be a least some minimal level of all three performance determinants in order for adequate levels of performance to be manifested. For a recent review of general theories of job performance see Tubré and colleagues (2006).

*Team Performance*

Like all measurement, valid measures of team performance must be based on theory (Salas, Burke, Fowlkes, & Priest, 2003). However, there is little consensus on definitions of team performance and team effectiveness (Kendall & Salas, 2004; MacBryde & Mendibill, 2003), which in turn complicates the process of creating general measures of team performance and effectiveness that are firmly rooted in theory. The issues involved with defining team performance should not come as a surprise given the struggle to define performance on the individual level (Avery & Murphy, 1998; Campbell, McCloy, Oppler, & Sager, 1993; Motowildo, 2003). The plethora of models of teamwork makes creating a highly generalizable framework for team performance measurement difficult. Instead, performance measurement should be customized to the context and grounded in a careful consideration of what is appropriate for a particular team (Kendall & Salas, 2004; MacBryde & Mendibill, 2003). Outcomes must always be defined in a manner that is relevant to the situational factors of the team being studied. Factors to consider are team type, team goals and objectives, and assessment objectives (Kendall & Salas, 2004).

In this vein, frameworks for developing team performance measures suited to a particular context have been formulated (e.g. Cannon-Bowers & Salas, 1997; MacBryde & Mendibill, 2003; Smith-Jentsch, Johnston, & Payne, 1998). According to Smith-Jentsch et al.'s (1998) framework, there are four dimensions of team performance that should be addressed by a measurement system: team processes, individual processes, team outcomes, and individual outcomes. Individual level measures need to be included in team measurement in order to identify and assist the team members that need help

developing their particular taskwork skills. In addition to team outcomes, team processes are an essential component of a measurement system (Canon-Bower & Salas, 1997). Four major team processes that have been empirically related to performance outcomes and that should be considered for inclusion in the measurement system are information exchange, communication, supporting behavior, and team leadership (Smith-Jentsch et al., 1998). However, it is recognized that even though there are some processes that are likely essential to all team performance outcomes, certain teams operating in specific contexts will probably require the development of unique measures to capture idiosyncratic or distinctive team objectives and patterns of interaction (Kendall & Salas, 2004). Therefore it is necessary to clearly define the team processes of interest in any one particular training context before attempting to formalize a measurement system. The rest of this section is dedicated to a review of the major conceptualizations of theoretical work with implications for the measurement of team processes. Specifically, the topics of IPO models, shared mental models, team adaptability, the ‘big five’ of teamwork, and macro-cognition/team cognition are discussed.

#### *Input, Process, Output (IPO) Models*

The number of theories and frameworks of team effectiveness and performance is vast—each theory or framework representing different research threads contributing to the larger picture of understanding teams (Salas, Stagl, & Burke, 2005). However, from this large and diverse field of team research emerges the central conceptual theme of current team research—the IPO model. Models and theories of this class seek to characterize the relationship between input variables such as individual and team characteristics, process variables such as shared cognition, leadership, communication,

coordination, decision-making, and back-up behavior, and outcome variables such as performance outcomes, productivity, and satisfaction (e.g., Gersick, 1988; Hackman, 1983; Marks, Mathieu, and Zaccaro, 2001; Salas, Dickinson, Converse, & Tannenbaum, 1992; Salas et al., 2005). By illuminating the relationships between input, process and output variables, these models and theories characterize teamwork as dynamic and multi-dimensional as well as highlighting the importance of team processes in effective teamwork (Guzzo & Dickson, 1996).

IPO models and theories in the literature have been systematically evaluated and an integrative framework has been produced (Tannenbaum, Beard, & Salas, 1992). This framework, illustrated in Figure 1, identifies four classes of input variables (i.e. work structure, individual characteristics, team characteristics, and task characteristics). The input variables relate to each other as well as to the team processes occurring over time. Team processes influence performance outcomes which are in turn looped back to team input variables in the form of feedback. Explicit in the Tannenbaum et al. framework is the importance of training in moderating the relationships between input and process variables as well as the relationship between process and outcome variables.

-----  
Insert Figure 2 about here.  
-----

### *Shared Mental Models*

Definitions of mental models vary but generally they can be thought of as knowledge structures involved in the integration of information and the comprehension of a given phenomenon (Johnson-Laird, 1983). Subsequently, a mental model that is

shared is a representation of information held by more than one person—team members. A shared mental model is an organized knowledge structure that facilitates interactions between team members (Klimoski & Mohammed, 1994). This “sharedness” of information models facilitates the effective coordination of action as well as promoting a similar method for processing new information within the team. These knowledge structures shared by individuals support the team members’ development of accurate causal explanations of their environment and tasks as well as accurate projections of future states of the environment and tasks. Consequently, shared mental models allow the team members to better adapt to the environment comprising task demands and their fellow team members (Cannon-Bowers, Salas, & Converse, 1993). This increased ability gained from shared mental models to better diagnose the environment and project future states has been linked with better decision making in teams (Stout, Cannon-Bowers, & Salas, 1996), in particular, their ability to recognize causal relationships in the external environment, make sound inferences, and create better explanations (Cannon-Bowers et al., 1993; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000).

The boon of shared mental models is especially significant in conditions that complicate communication between team members (e.g. extremely high workload, physical distribution, time pressure, stress). An increasing characteristic of teamwork is that team members are distributed in space and rely on various communication technologies to interact with each other. In this situation, shared mental models become crucial because they allow team members to predict the needs (e.g. information, material resources, present workload) of their teammates (Mathieu et al., 2000). Shared mental

models account for fluid, adaptive, and implicit coordination often observed in highly effective teams (Mohammed & Dummville, 2001).

### *Adaptability*

Historically, there has been a paucity of research that considered the temporal dynamics of team processes and performance (see Dyer, 1984) resulting in a lack of formalized understanding of how teams perform over time. However, this omission is being remedied by more recent models (e.g. Gersick, 1988; Morgan, Salas, & Glickman, 2001). Adhering to a multidisciplinary, multiphasic, multilevel perspective grounded in theory from cognitive, human factors, and industrial/organizational psychology, Burke et al. (in press) recently proposed a theory of team adaptation. At the center of team adaptation is adaptive team performance—defined as “an emergent phenomenon which compiles over time from the unfolding of a recursive cycle whereby one or more team members utilize their resources to functionally change current cognitive or behavioral goal directed action or structures to meet expected or unexpected demands” (Burke et al., in press, p. 8). The team adaptation model is illustrated in Figure 2.

-----  
Insert Figure 3 about here.  
-----

Burke and colleagues used an IPO approach to delineate a series of phases that produce adaptive team performance. Phase 1 consists of situation assessment comprising a cue recognition process and an ascription of meaning to cue patterns in the environment. Phase 2 is characterized by plan formulation, the phase in which the team develops and decides upon a course of action. Phase 3 is plan execution, which relies on

team coordination mechanisms. The final phase consisting of team learning, phase 4, arises from an evaluation of the team's performance. Knowledge of this evaluation will change how the team carries out the earlier phases of the model on future passes through the adaptive cycle. Therefore, as teams progress through these phases, emergent affective and cognitive states develop which are then called upon by team members in future performance episodes. In this way, Burke et al. characterize team adaptation as an episodic and recursive process.

This theory of team adaptation has significant ramifications for performance measurement in simulation. In a manner similar to traditional team performance, adaptive team performance can be viewed as a property of the team or as a configural construct (Kozlowski, Gully, Nason, & Smith, 1999). Therefore, depending on the organizational context, task-flow, interdependencies and other contextual variables, adaptive team performance will emerge in different ways (Klein & Kozlowski, 2000). Burke et al. (in press) propose that the ideal way to capture adaptive team performance is through a patterned emergence model. To capture adaptive team performance, metrics must be dynamic (i.e., capture performance over time) as well as context sensitive (i.e. they must be able to represent and be responsive to the changing nature of operational environment).

#### *The 'Big Five' of Teamwork*

The 'big five' of teamwork, illustrated in Figure 3, is an effort to unify the fragmented literature on teamwork and team effectiveness into a manageable set of features that can be used in practical contexts (Salas, Sims, & Burke, 2005). Salas and colleagues propose that the 'big five' components of teamwork—leadership, adaptability,

mutual performance monitoring, back up behavior, and team orientation—are important in all team contexts. In addition to these five factors, three coordination mechanisms—shared mental models, closed loop communication, and trust—are necessary for effective teamwork. Shared mental models and adaptability have been reviewed above; the remaining components of the ‘big five’ of teamwork are discussed below.

-----  
Insert Figure 4 about here.  
-----

*Leadership.* It is known that leaders have a considerable impact on team and organizational effectiveness (e.g. Zaccaro, Rittman & Marks, 2001). Increasingly, a functional perspective to leadership has been adopted (Fleishman, Mumford, Zaccaro, Levin, Korotkin & Hein, 1991; Hackman, 2002; Zaccaro et al., 2001), a perspective rooted in and understanding of leadership as “social problem solving that promotes coordinated, adaptive team performance by facilitating goal definition and attainment (Salas, Burke & Stagl, 2004, p. 343).” Functional leadership can be decomposed into four types of responses leaders can engage in to react to social problems: information search and structuring, information use in problem solving, managing personnel resources, and managing material resources. Increasing complexity in the task and social environments demands increased levels of adaptive reactions by the team and the team leader. Recently, the concept of shared leadership has been explored as a means to increase the adaptive nature of team leadership and thereby increase team performance.

In relation to traditional vertical leadership with more formal and static roles, shared leadership can be defined as “the transference of the leadership function among

team members in order to take advantage of member strengths (e.g., knowledge, skills, attitudes, perspectives, contacts, and time available) as dictated by either environmental demands or the development stage of the team” (Burke, Fiore, & Salas, 2004, p. 105). Shared leadership has been shown to be more effective than more traditional forms of leadership (Pearce & Sims, 2002). This increased performance has been attributed to shared leadership enabling the team members to shift leadership functions between one another in a manner that affords maximal exploitation of individual level knowledge and skills. Formal hierarchical leadership is not replaced by shared leadership; it is in fact the formal leadership that creates the team climate and structure that support the sharing of leadership within a team. The degree to which leadership functions can be relocated within a team—a coordination process—determines the effectiveness of this type of leadership.

*Team orientation.* Team orientation has been defined in a way that extends past a mere preference for working with team mates; it includes the tendency to coordinate, evaluate, and use the task inputs of other teammates, patterns of behavior that improve team and individual performance (Driskell & Salas, 1992). Empirical research findings have shown that when people experience stress their attention narrows as they focus inward; individuals become less willing to accept feedback/input from others as they tightly focus their efforts on their individual task (Driskell & Salas, 1991; Kleinman & Serfaty, 1989). Feedback and input from others is not only critical for obtaining optimal levels of performance in many situations, but it is vital to a team’s ability to learn and adapt within complex environments as well. By providing motivation and a higher baseline tendency to work as a team, team orientation may be able to partially ameliorate

the effects of stress within teams. Decision making performance is also positively influenced by team orientation (Driskell & Salas, 1992); this is an essential skill for teams that need to exhibit adaptive performance.

*Mutual performance monitoring.* Mutual performance monitoring is the ability to, “keep track of fellow team member’s work while carrying out their own...to ensure that everything is running as expected and ... to ensure that they are following procedures correctly...” (McIntyre & Salas, 1995, p. 23). Mutual performance monitoring can have negative effects on performance if it is perceived as a means of keeping tabs on team members or team members trying to protect themselves from responsibility for errors. That is why team attitudes must be accepting of mutual performance monitoring and the team culture must make provisions for the practice as a valid way of elevating levels of team performance. In addition to appropriate team attitudes, certain cognitive factors in the team are required for effective mutual performance monitoring. Specifically, team members must share mental models of the task, team, and equipment in order to correctly recognize abnormalities in their fellow team members’ performance as well as the needs of their teammates while engaged in mutual performance monitoring. Essentially then, the team members must know *what should* be happening before observing *what actually is* happening can yield fruitful information and boost individual and team performance.

*Back-up behavior.* Mutual performance monitoring is a necessary but not sufficient condition for developing and maintaining high levels of team performance. In order to capitalize on the mutual performance monitoring behaviors, team members must engage in back-up behavior. Back-up behavior is, “the discretionary provision of resources and task-related effort to another... (when) there is recognition by potential

back-up providers that there is a workload distribution problem in their team” (Porter, Hollenbeck, Ilgen, Ellis, West, & Moon, 2003, pp. 391-392). Back-up behavior is essential for effective team performance in three fundamental ways: (1) it affords the opportunity for team members to provide feedback to one another to improve performance, (2) it allows team members to provide assistance to one another during task performance, and (3) it allows for an adaptive capacity in teams such that the team can readjust its’ strategies by dynamically shifting workload between team members when an overload is detected in any one specific team member (Marks, Zaccaro, & Mathieu, 2000). Back-up behavior takes the form of either verbal or physical aid.

*Coordinating mechanisms.* The five core components of teamwork function by means of three core coordination mechanisms: shared mental models (discussed above), closed-loop communication, and mutual trust. Communication can generally be defined as the “exchange of information between a sender and a receiver” (McIntyre & Salas, 1995, p. 25). Communication processes bind the team’s work together and has been described as the glue of teamwork. Needed information is distributed throughout the team by means of communication. Similarly, shared cognitive structures are developed and their accuracy and completeness maintained through communication; this is necessary for creating common ground and adaptive performance. Research findings show that effective teams adapt their communication strategies by alternating between implicit and explicit communication. When a high performing team engages in explicit communication, that team will most frequently engage in a particular type of explicit communication – closed loop communication. Closed loop communication is defined by the following three dimensions: 1) a message being initiated by the sender, 2) the

message being received, interpreted, and acknowledged by the intended receiver, and 3) a follow up by the sender ensuring that the message was received and appropriately interpreted (McIntyre & Salas, 1995). The closed loop communication process is the most effective method of explicit communication because it ensures that there was no distortion of the message by the communication process, that is, closed loop communication confirms that the message was interpreted by the receiver in the manner intended by the sender.

In addition to shared mental models and closed-loop communication, mutual trust is an essential coordinating mechanism according the 'big five' of teamwork. Mutual trust has been defined in team contexts as "the shared perception...that individuals in the team will perform particular actions important to its members and...will recognize and protect the rights and interests of all the team members engaged in their joint endeavor" (Weber, 2002, p. 205). Without mutual trust, team members will waste attentional and communication resources checking to ensure that other team members are performing their tasks appropriately (Cooper & Sawaf, 1996). With mutual trust, team members are free to spend those resources of time and effort collaborating and working on individual tasks. Additionally, trust is implicated in a number of team processes and outcomes such as willingness to disseminate information (Jones & George, 1998), member participation and contribution, outcome quality, and member retention (Bandow, 2001).

#### *Macro-Cognition/Shared Cognition*

Just as macro-cognitive aspects of individual HIP have been emphasized in recent research, the growing tendency to conceptualize groups and teams as information processing systems (Hinsz, Tindale, & Vollrath, 1997) is accompanied by an

investigation of team-level macro-cognitive processes, team cognition in real world settings. The rise in importance of teams for addressing complex work environments is paralleled by an increase in the amount of cognitive tasks allocated to teams. For example, teams are formed to assess situations, make decisions, design and monitor plans, and gather knowledge. Therefore, understanding and measuring cognition at the team level is a vital aspect of understanding team performance and effectiveness. However, lack of clear definitions of the constructs involved with team cognition, the measurement purpose and the influence of the task environment have often been cited as impediments to the development of appropriate measurement methods for team cognition (e.g. Mohammed, Klimoski, & Rentsch, 2000). This section begins with a general discussion of the state of team cognition measurement; specifically, issues revolving around the level of analysis and aggregation of individual data to team level representations. This section concludes with a more detailed discussion of several classes of knowledge states and cognitive processes generally classified under the heading of shared or team cognition.

Team cognition can be characterized as an explanation of the interaction of and dependencies between intraindividual and interindividual level processes (Fiore & Schooler, 2004). Therefore, team cognition is a multilevel construct and measurement methodologies must address this aspect of the nature of team cognition in order to create valid and reliable measures. In the bulk of the research to date, there are three categories of measures used to assess team cognition: elicitation methods, team metric methods, and aggregation methods (Cooke, Salas, Cannon-Bowers, & Stout, 2000). All three of these methods share the underlying assumption that team knowledge is a collection of

individual level knowledge. The three categories of techniques also build upon one another such that each successive category is designed to allow for an extrapolation of individual level data to data representative of team level cognition.

First, the knowledge elicitation approach focuses on extracting a representation of domain-specific mental model like knowledge, usually from an expert in the domain. This is usually accomplished with such techniques as observation, interview, mapping techniques, and protocol analysis. In general, the knowledge elicitation category of team cognition measurement has been applied to relatively long lasting and stable knowledge structures at the individual level. That is, these types of techniques assess mental models that are assumed to change over time, but at a relatively low rate. These techniques have not been widely applied to more dynamic, rapidly changing knowledge structures such as team situation models. However, there are notable exceptions to this generalization. Specifically, knowledge elicitation techniques have been applied to the assessment of team situational assessment (Cooke, Stout, & Salas, 2001; 1997).

The second category of measures is team metrics. This category of measures arises to fill the need of quantifying information gathered through the knowledge elicitation methods. Much of the data derived from knowledge elicitation techniques is in the form of graphical representations, and therefore difficult to analyze in several dimensions: 1) in terms of accuracy of the knowledge structure, 2) comparisons between individual team members (e.g. similarity or overlap measures), or 3) aggregations of individual level data to the team level. Metrics can be calculated from the knowledge elicitation data in terms of similarity (e.g. a percentage score for overlap in individual

knowledge structures), or accuracy (e.g. a percentage score for correct knowledge), as well as metrics that account for team heterogeneity and distribution of knowledge.

Lastly, team metrics need to be effectively translated into team level measures of cognition. This is done through various aggregation methods such as averaging the individual scores or, taking the median, maximum or minimum value. There is no clear method to aggregate individual level data to team level data in this manner as in some instances taking the mean value of a team metric will be misleading, particularly when the variance in individual scores is high. Cooke et al. (2000) recommend using individual metrics to categorize teams versus using an aggregate team measure. In this way, relationships between various configurations of teams can be explored, a capability not easily realized with aggregate data.

Researchers are currently exploring options to measurement techniques that attempt to aggregate individual level data to team level representations of cognition as some conceptualizations of team cognition imply that the aggregation based methods are not tapping appropriate types of knowledge and cognition. For example, Klimoski and Mohammed (1994) describe team knowledge as an emergent phenomena resulting from the interaction of team members. This perspective places a greater emphasis on team processes and allows for the possibility that team knowledge is something different than the sum of the team's individual members' knowledge. Cooke, Salas, Kiekel, & Bell (2004) present a framework of team knowledge that elucidates the realm of knowledge (i.e. individual and relatively static structures) tapped by the traditional methods described above as well as the team knowledge that is left untapped (i.e. knowledge involving team processes such as communication, situation awareness, and coordination).

The rest of this section discusses several shared cognition constructs in more detail: transactive memory, team situation awareness, and metacognition.

*Transactive memory.* The construct of transactive memory is an expansion on most conceptualizations of SMM in that it speaks to the storage of information, not the interrelationships among the stored items. Wegner (1986) defines a transactive memory system as (1) a combination of the knowledge possessed by each individual and (2) a collective awareness of who knows what. Yoo (2001) argues that it is team members' meta-knowledge about who knows what in a team. Austin (2003) expands this definition by adding transactive memory consensus/agreement and accuracy as components. The transactive memory system is founded on the idea that, due to complexity, individuals know only part of what the team as a whole knows and this team knowledge is distributed unequally amongst members (Moreland, Argote, & Krishnan, 1986). Research has begun to indicate that transactive memory systems can enhance a team's performance, particularly when the task is complex and requires a considerable contribution of knowledge from individual team members (Faraj & Sproull, 2000; Lewis, 2000; Moreland, 1999).

*Team situation awareness.* Situation awareness (SA) has been defined as the 'reception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future' (Endsley, 1995, p. 36). SA varies on many dimensions, but accuracy and completeness are of greatest interest. The accuracy of an individual's SA is the degree to which a component of knowledge in an individual's SA represents the external environmental reality. The completeness of an individual's SA is the degree to which knowledge in

individual's SA captures all or only some of critical environmental factors. As represented in the above definition, there are three levels of SA: perception, comprehension, and projection. Accuracy and completeness apply to each level of SA.

Although often described at an individual level of analysis researchers have begun to conceptualize team situation awareness (TSA). For example, Salas, Prince, Baker and Shrestha (1995) suggest that TSA is more than the sum of the individual members' SA and includes team process behaviors as well. Stout, Cannon-Bowers and Salas (1996) argue that the work on TSA has a number of implications: (1) existing mental models shape individual SA; (2) shared mental models seem to be critical to team performance because they allow members to form adequate explications and expectations of task and team actions; (3) TSA enhances team performance. After conducting critical incidents reviews with pilots, Prince and Salas (1997) identified four major actions important for TSA in commercial pilots: identifying problems or potential problems, demonstrating knowledge of the actions of others, keeping up with flight details, and verbalizing actions and intentions.

*Metacognition.* Flavell (1979) conceptualized metacognition as deliberate, goal directed, and future orientated mental behaviors that are directed toward accomplishing a task. Kluwe (1987) extended this work and identified two attributes of metacognitive activity. First, the thinking subject has some knowledge of their own thinking and the thinking of others. Second, the attribute consists of active monitoring and regulation of one's own thinking, where one is acting as the causal agent of their own thinking. Consistent with Kluwe's work, Mazzoni and Nelson (1998) describe two primary functions of the metacognitive system: control and monitoring. Metacognitive control

refers to activities taken by individuals to influence the course of their own thinking, including initiating, monitoring, and prioritizing mental activities. Metacognitive monitoring consists of updating the meta level model of the situation with information from the object level. This model can then be evaluated with respect to the goal the system is attempting to achieve.

Research investigating team members' higher order cognitive structures has focused on illuminating the role of metacognition (Cohen, Freeman & Wolf, 1996; Kozlowski, 1998; Lord & Emrich, 2001; Sternberg, 1998). It has been asserted that "at the team level, metacognitive and self regulatory processes must extend beyond the self" (Kozlowski, 1998, p. 120). This line of research suggests that teams can become self learning, adaptive systems when they have developed metacognitive skills at the team level. Accumulating research findings suggest that team member metacognition can be accurately measured and that obtained scores are related to important team outcomes such as leader performance (Koslowski, Gully, Smith, Nason & Brown, 1995; Marshall-Mies, Fleishman, Martin, Zaccaro, Baughman, & McGee, 2000), team performance (Marsh, Kiechel, Boyce & Zaccaro, 2001), as well as tacit knowledge and social competencies (Banks, Bader, Fleming, Zaccaro, & Barber, 2001).

In sum, a large array of theoretical frameworks, models, and approaches have been used to conceptualize various aspects of human performance at the individual and team levels. This is not surprising given that human performance in general is multifaceted and highly complex process. A complex approach to measurement must be taken to accurately capture a complex phenomenon. Multiple measurements rooted in the various theoretical approaches discussed above can help to provide a deep and detailed

view of human performance in training environments. To this end, measurement methodologies are discussed in the next section.

### **What Methodologies are being used?**

There are a variety of methods in use to measure performance at the individual and team levels. These range from quantitative to qualitative and from subjective (e.g. self-report measures) to objective (e.g. task outcomes). A third dimension by which to categorize performance measures is the level at which they capture performance (e.g. individual or team). This aspect of performance measurement is less solid than the issue of whether or not the measure is qualitative or quantitative, subjective or objective. With due care, some of the same methodologies can and have been used to measure different levels of performance. However, the nature of the construct greatly impacts how that construct is measured in a multi-level manner. Two main 'pure-type' distinctions involved in developing measures of team level performance based on individual performance are compilation and composition. Composition describes a process whereby a lower level (e.g. individual) emerge upward to the team level but remain essentially the same. Conversely, compilation is the process whereby similar but distinctively different lower-level properties combine into a higher-level (e.g. team) property that is related to but different from its diverse lower-level constituent parts (Kozlowski & Klein, 2000).

Another important factor in performance measures is the implicit conceptualization of performance that the measure embodies; that is, does the measure originate from the perspective that performance is a process or an outcome? Campbell argues that "performance is not the consequence or result of action; it is the action itself (1990, p. 174)." This sentiment is echoed in the team research, specifically, that all too

often team performance is conceptualized as the products or outcomes of actions that extend over time and not the actual processes themselves (Kozlowski & Bell, 2003). However, this approach does not dismiss the importance of performance outcomes as they are often what is ultimately of interest. Essentially, if performance is conceptualized and measured as a process and this aids in the performance measurement goals of diagnosis and intervention, then there is the underlying assumption that by improving process, outcomes will also benefit.

The rest of this section is dedicated to an overview of some of the performance measures in use. Measures are grouped into two major classes—qualitative and quantitative. Other aspects of the measure such as level of analysis and specifics about what the measure captures are discussed for each measure. Table 1 summarizes the measures discussed in this section. It details the type of data the measure relies on, the level of analysis, whether the measure captures processes or outcomes, and key sources.

#### *Qualitative*

Many of the techniques classified as qualitative are used to develop an understanding of what performance consists of in a given domain. That is, qualitative measurement techniques can be used in the development of performance measurement systems which are subsequently used to capture performance during SBT. Qualitative measurement techniques are better suited for building an understanding of the processes that drive effective performance in the real world (e.g., the goal of SBT). More quantitative measures of performance can then be developed which capture these processes and enable feedback, remediation, and assessment during SBT.

#### *Protocol Analysis*

In support of measuring internal individual and team processes, protocol analysis is a method for using verbal reports as indicators of cognition (Ericsson & Simon, 1993). This is an increasingly valuable tool for training and simulation as the tasks trained become more complex and cognitively based. There are two general categories of protocol analysis: on-line and off-line (Shadbolt, 2005). On-line variants involve the participant verbalizing his/her thoughts concurrently with task performance. In off-line protocol analysis techniques, the participant views a video recording of his/her task performance session and verbalizes his/her thoughts. Generally, on-line analysis techniques are thought of as more naturalistic while off-line techniques are assumed to interfere less with task performance. An additional related technique is the cognitive walkthrough which involves participants verbalizing how they would perform a task, but without actually doing the task (Bainbridge & Sanderson, 2005).

Protocol analysis is a three stage process leading from research or design questions to conclusions (Bainbridge & Sanderson, 2005). First, the data is collected; that is, the verbalizations are recorded and transcribed. At this stage, it is important to capture as much of the task context as possible so that ambiguities arising in the analysis at a later time can be resolved. For this reason, video recording is preferred to just audio. It is also good practice to have the participant review the transcripts if possible. Second, the protocol data is prepared for analysis. The researcher identifies a general structure of related segments in the protocol. This resembles a task analysis in that the utterances are grouped into hierarchical sections of activity. Next, the analyst identifies meaningful units that appear to represent separate mental processes. From this decomposition of the protocol, a structure of mental activities can be inferred, formal descriptive languages can

be developed to aid in statistical analysis and abstraction of the protocol, and inferences about processes not represented in the protocol can be made (Bainbridge & Sanderson, 2005). The third and final stage of verbal protocol analysis is the examination of explicit and implicit content achieved through various methods of content analysis, sequential analysis, and modeling techniques.

#### *Critical Incident Technique*

Since its development in the Aviation Psychology Program of the US Air Force during World War II, the Critical Incident Technique has been applied in a diverse range of research domains (see the APA Critical Incident Technique Bibliography, Fivara, 1980). Flanagan (1954) describes the critical incident technique as a “set of procedures for collecting direct observations of human behavior in such a way as to facilitate their potential usefulness in solving practical problems (p. 327).” An incident is sufficiently complete human behavior that, in and of itself, allows for inferences about that agent. An incident is critical if it occurs in a situation that: 1) the intent of the agent is sufficiently clear to the observer, and 2) there is little doubt about the effects of the behavior, as the observer can clearly identify the consequences of the action. The technique does not comprise a hard and fast set of immutable rules, but a flexible set of principles that must be adapted to specific situational characteristics. There are three core characteristics of the technique: observers are only required to make the simplest of judgments, only qualified observers participate, and an a priori agreed upon purpose of the research is used to guide the evaluation of observations. These elements of the Critical Incident Technique serve to minimize the amount of observer bias and interpretations in the

observations. The technique can be used to measure typical performance and proficiency.

Additionally, the Critical Incident Technique is heavily relied upon to develop scenarios and behavioral triggers embedded in scenario-based training and therefore a valuable tool in developing event-based measurement systems (e.g. Fowlkes, Lane, Salas, Franz, & Oser, 1994; Morgan, Lassiter, & Salas, 1989). A more recent adaptation of the Critical Incident Technique is the Critical Decision Method which is concerned with determining factors influencing situation assessment and decision making (Klein, Calderwood, & MacGregor, 1989). In many ways the Critical Decision Method is a hybrid between the Critical Incident Technique and retrospective protocol analysis in that it focuses on probed interviews with individuals concerning performance on non-routine events. The Critical Decision Method is also related to Cognitive Task Analysis (Hoffman, Crandall, & Shadbolt, 1998).

### *Concept Mapping*

An important technique in measuring team knowledge, a central component of team cognition (Cooke, Salas, Kiekel, & Bell, 2004), is concept mapping or cognitive mapping. This knowledge elicitation approach measures the structure, content, and robustness of the knowledge of individuals and teams and generates a graphical representation of how cognitive concepts are related to one another for any one person. This technique derives much of its popular support from its treatment and analysis of meaning as a relation between knowledge elements. Because these discrete elements of cognitive content can be related in numerous ways (e.g. causality, resemblance, continuity, proximity, etc.; Weick & Bougon, 1986) many different types of concept

maps exist; however, maps representing causality between elements are generally the most widely used (Mohammed, Klimoski, & Rentsch, 2000) as causality is considered the most important relation between knowledge elements (Gray, Bougon, & Donnellon, 1985).

In addition to the existence of different types of concept maps based on the relations represented in them, there are several different methods for generating the maps. First, data is gathered from the participants in one of two general approaches: 1) questionnaires and interviews (Mohammed et al., 2000), or 2) through analysis of pre-existing data (Carley, 1997). Second, in relation to the questionnaire and interview techniques, there are several variations based on whether the content, structure, both or neither are provided to the participants (Salas, Priest, & Burke, 2005). In one instance, the research gives the participant a set of concepts, which must be placed within a pre-defined structure. Another case involves the participant being given only the structure and is then required to generate the content on his/her own. Lastly, participants may be required to generate both the structure and the content.

### *Quantitative*

The measurement approaches discussed above are primarily used to develop an understanding of the criterion; they help to answer the ‘what to measure’ question. After this understanding of what drives performance outcomes in a domain, quantitative measures that sample this criterion space can be developed. This section discusses the major types of quantitative measurement tools available for capturing performance during SBT.

### *Behaviorally-Anchored Rating Scales (BARS)*

BARS is a widely used technique first developed by Smith and Kendal (1963) and a very popular way of promoting accurate and consistent observer ratings when rating and classifying individual and team behaviors. BARS consist of a series of brief descriptions that characterize different levels of performance. Each description is associated with a specific scale value and the behavioral examples represent episodes of excellent, acceptable, and poor performance. These descriptions of behaviors and anchor points are frequently generated from SME accounts using the Critical Incident Technique (Flanagan, 1954; Fowlkes et al., 1994; Morgan et al., 1989) which can be used to characterize the upper (i.e. high quality performance) and the lower (i.e. poor performance) bounds or anchor points of the BARS. Because BARS include descriptions of specific behaviors, one concern in the use of BARS involves the tendency for observers to focus only on the behaviors that are a part of the checklist (Murphy & Constans, 1987; Murphy & Pardaffy, 1989).

#### *Behavioral Observation Scales (BOS)*

Like BARS, the purpose of BOS is to facilitate accurate observer ratings of team performance; however, BOS assesses typical performance, unlike BARS that assesses isolated performance. BOS accomplish this through the use of a Likert scale that observers use to record the frequency with which teams exhibit certain behaviors and engage in targeted processes. The team behaviors and processes must first be conceptually defined and examples generated from SMEs in a manner akin to that used in BARS. These examples are then distilled into cogent and concise statements. During the simulation or training exercise, the observer then issues a numerical frequency score to each of the process or behavior statements (e.g. ranging from 0 = never, to 9 = always).

Because this method of observational rating forces the raters to evaluate over a period of time and rely on their memories, it is important that observers be trained to avoid falling prey to known human memory contaminants to data such as primacy and recency effects (e.g. Baddeley & Hitch, 1993). Meta-analytic review has shown that it is possible to achieve and maintain high reliability and accuracy of frequency-based judgments as ratings despite some inherent biases (e.g., overestimating infrequent events and underestimating highly frequent events;

#### *Communication analysis*

As of now, team processes are best measured through communication analysis. Most frequently, communication has been measured in terms of simple frequency as this approach is the most straightforward and easy to implement. However, research attempting to link frequency of communication with team outcomes has produced mixed results (e.g. Foushee & Manos, 1981; Jentsch, Sellin-Wolters, Bowers, & Salas, 1995; Orasanu, 1990; Orasanu & Fischer, 1991). This has spurred the development of more sensitive methods for analyzing team communication.

Content analysis is a specific type of communication analysis that can be used in conjunction with other procedures such as protocol analysis. Its aim is to measure aspects of verbal exchange that related to team processes such as coordination. In one approach to content analysis, communication is coded with the use of a speech act typology (Foushee & Manos, 1981). The typology can be exhaustive or simplified to include only fundamental aspects of communication such as the initiating of communication and the response to communication with each of these two categories broken down to include types of communication (e.g. commands, questions, acknowledgements, etc.; Kanki,

Lozito, & Foushee, 1989). Patterns in the types of communication can then be analyzed by calculating matrices of initiating and responding speech as well as ratios of initiating and responding speech (Kanki et al., 1989).

Recently emerging techniques include attempts to automate the communication analysis process through the use of Latent Semantic Analysis (LSA; Keikel, Cooke, Foltz, & Shope, 2001). Another approach, Exploratory Sequential Data Analysis (ESDA; Sanderson & Fisher, 1994), has been used to analyze temporal patterns in team communications (Bowers, Jentsch, Salas, & Braun, 1998). Bowers et al. coded each utterance into categories (e.g. action, response) and then analyzed the sequence in which the statements were made. This line of research linked a specific sequence of communication, closed-loop communication, to higher levels of team performance. Additionally, Social Network Analysis has been used to create measures of team mental model congruity from voice (Graham, Schneider, Bauer, Bessiere, & Gonzalez, 2004) and text communications (Carley, 1997).

#### *Event-based measurement*

Targeted Acceptable Responses to Generated Events (TARGETs) (Dwyer, Fowlkes, Oser, Salas, & Lane, 1997; Fowlkes, Lane, Salas, Franz, & Oser, 1994) is an example of event-based performance measurement. This is a structured observation methodology in which teams are exposed to scenarios consisting of contextually relevant exercises or tasks created by the researcher/trainer. These scenarios contain cues for team members to exhibit behaviors that have been identified as important for that particular team. In addition to defining the tasks considered desirable to observe, the researcher/trainer must determine what an acceptable response to the scenario is a priori

by means such as subject matter expert interview, task analysis, or investigation of standard operating procedures. Acceptable responses are determined in advance so that the observer can have a checklist at the time of the observation. This significantly adds to the reliability of the observational rating. Event-based measurement such as the TARGETs methodology provides the opportunity to observe team behaviors that have a low frequency of occurrence in the normal operation of the team and therefore are difficult to observe in a purely naturalistic setting (Fowlkes, Dwyer, Oser, & Salas, 1998).

### *Pathfinder*

Pathfinder is a technique related to concept mapping in the sense that both examine structural properties of knowledge representation; however, the techniques differ in terms of the methodologies used to accomplish this task (Salas et al., 2005). Essentially, Pathfinder represents knowledge in a network structure with the concept as the node and the degree of relatedness between concepts as the relations between nodes (Mohammed et al., 2000). Pathfinder then algorithmically transforms the network measures into a psychologically scaled proximity data and a representation of the underlying structure of knowledge elements (Schvaneveldt, 1990). Constructs that are rated as highly similar are visually represented with direct links; those with more tenuous relationships are represented with indirect links, and no link is present between concepts with no relationship. In addition to the visual representation of knowledge structure, Pathfinder generates a quantitative index of coherence (Salas et al., 2005). To generate this data, the researcher must first identify a list of concepts to be presented to the participants. This list is most often derived from task analysis procedures. The concepts

are then presented to the participants who make a pairwise rating of each possible pairing of concepts.

### *Self-report measures*

Questionnaires are a popular method for assessing individual and team level cognitive and affective states and processes. Many measurement systems rely on the use of data collected at the individual level through self-report measures which is then analyzed and translated into team level characteristics (e.g. team orientation, Driskell & Salas, 1992; collective efficacy, Tagger & Seijts, 2003; team cohesion, Gully, Devine, & Whiney, 1995). These techniques involving aggregation or translation between levels require careful analysis of the constructs and the measures involved. When translating measures from one level to another it is important to determine whether the underlying construct should be characterized as global, shared or configural (Kozlowski & Klein, 2000). Constructs with global properties only exist on the team level (i.e. team size and structure). Constructs with shared properties are common to each member of the team (e.g. collective efficacy, trust). Constructs with configural properties are a composite of different types of individual level properties into a team level construct.

### *Automated Performance Recording and Measurement*

Many process variables such as communication and behavior can be automatically and continuously monitored and recorded during training sessions in a simulation. These automated measures have several advantages over traditional measurement including being less invasive to the performance episode and containing fewer inherent biases. Automated performance recording and measurement is limited to direct and overt behavioral responses and is therefore ineffective, by itself, for measuring

cognitive processes; however, types of communication analysis (e.g. sequential analysis, latent semantic analysis) are being developed to address this issue (e.g. Keikel et al., 2001). The cost of implementing automated performance recording and measurement systems can be prohibitive for many measurement purposes.

To summarize, performance measurement systems for simulation should be rooted in theoretical models of performance. Different theories will be relevant to different simulation objectives and therefore different measurement requirements and techniques should be applied. The simulation purpose and context should drive the selection of the theories used in the development of performance measurement systems. The previous section reviewed a sample of techniques and tools available for measurement in simulation. The following section applies these ideas to the available literature on performance measurement in simulation training and distills a set of best practices for designing and implementing measurement systems.

### **What are the Best Practices in Current Simulation Based Training Systems?**

Given the breadth of theories driving performance measurement and the variety of tools used to assess performance, the process of bringing these components together to design and implement a performance measurement system for use in a simulation training environment can be filled with uncertainty. Therefore, this section reviews the literature on performance measurement systems to distill a set of best practices. There are several issues involved with an attempt to assess the state of performance measurement systems in simulation and training from a review of the literature. First, the purpose of much of the published literature is experimental in nature; therefore, the measures collected and reported support the purpose of experimentation (e.g. comparing groups) rather than the

purpose of training. The reader does not know whether the measures reported in an experimental paper are the same as those used for that simulation in a standard training situation. Good practice dictates that the purpose of measurement should drive the development of measurement systems (Brannick & Prince, 1997), and therefore the measures reported for the purposes of experimental research are likely different than those used for the purposes of training. Full documentation of the measurement systems used for training simulations is not widely disseminated. Cannon-Bowers and Salas (1997) advanced a set of requirements for performance measurement in training systems. While these requirements deal specifically with team training, with few exceptions they are generally applicable at the individual level as well. In this section, the basic requirements of a performance measurement system are briefly discussed and used as a framework to organize the best practices distilled from the literature.

*Performance measurement systems must consider multiple levels of measurement*

The multi-dimensional and multi-level nature of performance has been a recurring theme in the previous sections. In the simplest case, performance can be measured at the individual and team level. Adding complexity, team performance can be measured at the team level or as an aggregate of individual level properties (i.e. the issue of compilation vs. composition discussed earlier). Therefore, a critical factor in designing a measurement system is choosing the appropriate level of analysis for the measurement system. The type of skill being trained and the type of feedback and remediation necessary to improve performance determine the appropriateness of the level of analysis for measurement.

*Best practice 1:* Performance measurement works best when it captures multiple dimensions of performance at appropriate levels of analysis.

Performance measurement and evaluation on multiple levels is a major issue in any situation involving more than one person and especially for joint collective operations simulation (Fleming, Barto, & Johnson, 1998). Fowlkes et al. (1999) describe the development of the Tactically Relevant Assessment of Combat Teams (TRACTs) performance measurement system for capturing performance data at multiple levels of analysis. The TRACTs system was implemented in a synthetic theatre of war environment and used an event-based approach to measurement for capturing interactions constituting performance on the team and sub-team levels. As discussed in the methodology section, event-based approaches such as TRACTs and TARGETs (Dwyer, Fowlkes, Oser, Salas, & Lane, 1997; Fowlkes, Lane, Salas, Franz, & Oser, 1994) ease the burden on observers because observers are aware of when and where targeted individual and team behaviors would occur. The a priori specification of targeted individual and team level competencies of interest simplifies the process of observing multilevel multidimensional performance.

Carolan et al. (2003) describe the development of a performance measurement system designed to capture performance in Distributed Mission Operations (DMO) training simulations. This system combines objective performance data captured by the simulation with observer ratings and information from trainees captured from post-training questionnaires and debriefs. By combining these various methods, the measurement system provides a robust basis for evaluation and remediation and by employing a common measurement framework the process of integrating the various methods is facilitated. The system creates a profile of both individual and team performance with multiple methods of measurement contributing to both levels of

analysis. As one of key purposes of performance measurement in SBT is feedback, and timeliness is key to the effectiveness of feedback, there must be a plan in place for systematically integrating all of this data into a form that is meaningful and meets the goals of measurement (Oser, Cannon-Bowers, Salas, & Dwyer, 1999).

*Best Practice 2:* Performance measurement works best when event-based techniques are used to capture data at multiple levels of analysis.

*Best Practice 3:* Performance measurement works best when multiple measures from various sources are captured.

*Best Practice 4:* Performance measurement works best when a systematic plan is in place to integrate data from multiple measures.

*Performance measurement systems must consider processes in addition to outcomes*

In a training environment, knowledge of performance outcomes does not provide enough information to diagnose performance deficiencies nor does it provide a basis for corrective feedback; therefore, measurement systems must capture process dimensions of performance as well as outcomes. Outcomes alone do not directly describe the reasons that a certain level of performance is achieved. Outcome performance measures can alert trainees and trainers that a problem is present, but they provide no direction as to correcting the problem. In this sense, measurements of process are more diagnostic. Additionally, processes are more descriptive of what is to be trained.

*Best practice 5:* Performance measurement works best when it captures the processes of performance.

Leddo, Zhang, and Pokomy (1998) report on the development of an automated performance measurement system embedded in an in-patient care simulator for medical simulation. The performance measurement system uses an expert model of the task to evaluate trainee performance behaviors during the training episode as well as trainee

responses to a post-training questionnaire presented automatically by the simulation. By combining these methods, the system is able to ensure that the trainee has developed proficiency with the task as well as conceptual understanding of the principles being trained.

Schreiber et al. (2003) discuss the development of a proof-of-concept for a Performance Effectiveness Tracking System (PETS) for Distributed Mission Operations (DMO) training simulations. A core characteristic of the PETS is the need to capture performance process as well as outcome measures at the individual and team levels. To accomplish this, the researchers created a tool that 'listens' to the data shared between various nodes in the networked simulation. The system captures multiple measures of performance by tracking the system state and trainee behavior in real-time.

*Best Practice 6:* Performance measurement works best when expert models of the task are used to evaluate performance processes and outcomes.

*Best Practice 7:* Performance measurement works best when the collection and transmission of objective measures is automated.

*Performance measurement systems must describe, evaluate, and diagnose performance*

By addressing the above issues (i.e. should the measurement system capture performance at the process or outcome phase, at the team or individual level), a performance measurement system is much more likely to meet its overall objectives. Generally, these objectives are the description, evaluation, and diagnosis of performance. The description of performance is conceptually simple, although in practice is often very difficult due to the complexity and dynamic nature of tasks. The evaluation of performance is the assessment of observed behaviors and judgments concerning their appropriateness. In order to make an evaluation of performance, the measured

performance must be compared to some criterion or standard for the desired level of performance (Oser et al., 1999). Diagnosis of performance is the determination of the causes of both effective and ineffective performance. The process of performance diagnosis involves developing a rich and detailed profile of performance that can be used to identify the presence or absence of competencies targeted for training (Salas, Rosen, Burke, Nicholson, & Howse, in press).

*Best practice 8:* Performance measurement works best when measures are descriptive of performance.

*Best practice 9:* Performance measurement works best when performance can be compared to standards for desired levels of performance.

*Best practice 10:* Performance measurement works best when it is diagnostic—when it provides insight into the causes of performance.

The expertise of the trainer is an essential component of many training simulations. However, the complexity of many training situations complicates the trainer's task of evaluating and diagnosing performance. Drewes and Gonzalez (1997) present a method for automatically monitoring, evaluating, and diagnosing trainee performance during simulation training. The system—called Template Based Reasoning (TBR)—uses artificial intelligence to assist in the interpretation of trainee behavior. A template consists of attributes that define the particular behavior being trained. The system monitors the trainee's current behavior and creates a measure of confidence as to how well the current behavior matches the template. This analysis can be fed into automated feedback to the trainee as well as being used by the trainer for evaluation and remediation. A similar approach was taken by Zachary and colleagues (1997) in the development of an embedded training system for shipboard systems. Their system used a cognitive architecture and real-time broad spectrum capture of trainee behaviors to

compare what the trainee *is* doing to what they *should* be doing. This process is used to create feedback to the trainee.

*Best Practice 11:* Performance measurement works best when it allows for performance evaluation to be partially or fully automated.

*Best Practice 12:* Performance measurement works best when it allows for performance diagnosis to be partially or fully automated.

Stacy et al. (2004) discuss the development of Human Performance Objects (HPOs), a software object used for collecting performance data in federated High Level Architecture simulation and training environments. HPOs enable the effective collection of performance data in distributed environments, and provide a clear method for substituting various measures of performance depending on the training needs by integrating performance measurement with the software architecture of the simulation itself. This characteristic of HPOs is extremely helpful for distributed mission training because performance measurement data is incorporated into the information shared by the various distributed simulation nodes.

*Best Practice 13:* Performance measurement works best when there is flexibility designed into embedded measures (different measures can be substituted).

Cooper and colleagues (2003) describe an embedded training system within military aircraft that supports realistic in-flight training and pre-mission rehearsal while on route to the actual mission. An important feature of this system involves mission playback. The system records all of the information available to a crew member; during a training session, a single crew member may have several different screens from which to select. After the training session, the trainer can choose the screen the trainee was

using at a particular moment or any of the other screens that were available to the trainee. This ability gives the trainer access to the entire set of information available to the trainee and assists in the diagnosis of effective and ineffective performance and allows for appropriate feedback and remediation.

*Best Practice 14:* Performance measurement works best when it captures a broad spectrum of measures and the context of performance.

Presently, observers are still necessary components of performance measurement systems, especially when evaluating and diagnosing complex performance. When observers are used, several steps need to be taken to ensure that quality data is collected. Observers must be trained to be highly reliably in terms of internal consistency as well as inter-rater reliability (Baker & Salas, 1992; 1997). As noted in the methodology section, many of the tools available to observers to aid in data collection (e.g., BARS, BOS) may introduce specific biases. Training with the tools to be used during observation can reduce the method variance and yield better measurement. In addition to training, protocols need to be developed to guide the observers' attention to salient aspects of performance and to improve the consistency and quality of data gathered during observation sessions (Dwyer, Fowlkes, Oser, Salas, & Lane, 1997; Fowlkes, Lane, Salas, Franz, & Oser, 1994).

*Best Practice 15:* Performance measurement works best when observers are trained to high levels of reliability.

*Best Practice 16:* Performance measurement works best when observers use protocols.

*Performance measurement systems must provide a basis for remediation*

Performance diagnosis feeds into remediation, the next requirement of a performance measurement system. In a simulation training environment, the performance measurement can be used for a number of purposes; however, the overriding purpose of measurement in simulation is to support the instructional process. Measures should inform the trainer on trainee progress, provide trainees with information of results and feedback on how to improve performance, and provide a basis for selecting future training.

*Best Practice 17:* Performance measurement works best when it supports learning.

A common form of feedback to trainees during training simulations in a military environment is the After Action Review (AAR). Traditionally, AAR systems have been add-ons to training simulations and have been criticized for resting on methodologically unsound ground (Kelly, 1999). In a brief review of the development of AAR systems in simulations, Kelly (1999) documents a trend in systems running from reporting basic outcomes (e.g. number of enemies destroyed, number of casualties, completing mission objectives) to providing more sophisticated analysis of the performance episode and tools to support changing processes involved in performance. Meliza and Brown (1996) document the Automated Training Analysis and Feedback System (ATAFS). This system was built to address the problem faced by trainers in creating AAR aids. Specifically, trainers lack the time to create effective AAR aids in a timely manner for training remediation. ATAFS was developed to ease the burden on training instructors by automating the performance measurement process during simulations as well as the construction of AAR aids for training remediation after performance episodes. The system automatically generates a bin of visual aids for AAR from performance data

automatically captured by the simulation as well as facilitating manual development of visual aids by the trainer.

*Best Practice 18:* Performance measurement works best when it allows for the automated and manual creation of AAR aids for training remediation and feedback.

Additionally, Frank et al. (2004) describe the use of an automated performance measurement system embedded within a distributed training simulation. In this environment the performance measurement system is used to automatically generate AARs for students. The AARs are used to scaffold the training, that is, adapt the training to the trainee's performance. This has proven to be a very effective method for delivering training without the presence of trainers. Chandler and Owens (2003) report on a similar approach using intelligent tutoring systems to facilitate performance-based coaching.

*Best Practice 18:* Performance measurement works best when it enables automated scaffolding and performance-based coaching.

Berger (1999) discusses the use of real-time corrective feedback in the training of truck drivers. In the SIGNAL MAN system, the trainer specifies an endpoint and path for the trainee to follow. Trainee input to the simulation is then evaluated by algorithms embedded in the training system that calculate a route based on present and desired (end) system states. A corrective feedback loop is created by displaying cues to the trainee based on the calculated route.

*Best Practice 19:* Performance measurement works best when it drives real-time corrective feedback.

Meliza and Paz (1996) discuss the need for training systems to integrate the feedback and control functions of a training system. In most training systems, trainers are given control over many situation or scenario variables within the simulation. By connecting the manipulations made to these variables in the simulation to the process of generating feedback (e.g. creating AARs), the workload on the trainer(s) can be greatly ameliorated. In this way, the simulation can be modified with any combination of available control variables, and the automated system measurements will provide appropriate feedback regardless of how the simulation unfolds.

*Best Practice 20:* Performance measurement works best when it is integrated with trainer controls and feedback generation.

### **Where Do We Go From Here?**

This chapter has highlighted a subset of the theories of human performance used to guide measurement in simulation based training as well as the methodologies commonly used to capture data. Additionally, the practice based literature has been reviewed to illustrate how theory and methods can best be applied in the development of performance measurement systems in simulation. We conclude with some comments on needs for continued research in the areas of theory, methods, and applications. First, although theories of human performance are plentiful, there is a strong need for more precise, yet ecologically valid, theories of human performance. All too often the robustness of implications a theory has for performance in complex environments is sacrificed for precision and vice versa. To progress, the development of performance measurement systems in simulation based training needs theories of human performance that can represent the complexities of interactions between trainees and the simulated

environments while simultaneously affording a detailed and precise understanding of the multiple dimensions of performance. Second, methods need to be developed that allow for the moment-to-moment capture of data that is necessary for performance diagnosis. It is necessary to measure the processes of performance to understand the causes of performance and therefore a focus must be placed on the capture of data over time.

Third, applications must be developed that partially or wholly automate performance measurement, evaluation, and diagnosis. There has been advancement in the automated analysis of communication content and flow (e.g. Kiekel, Cooke, Foltz, & Shope, 2001), as well as the capture of task performance over time, but this is just the beginning. Performance diagnosis requires the integration of multiple measures from multiple data sources and analyzed at multiple levels. To accomplish this in a systematic and repeatable manner requires development in the areas of theory and methods described above; however, the best practices presented in this paper provide examples where designers of performance measurement systems have applied the current theories and methods in a manner that yields maximum effectiveness of performance measurement in simulation based training.

### **Acknowledgement**

This work was supported by the NAVY PERSONNEL RESEARCH, STUDIES AND TECHNOLOGY DEPARTMENT (Janet Held) under the auspices of the U.S. Army Research Office Scientific Services Program administered by Battelle (Delivery Order 0665, Contract No. DAAD19-02-D-0001).

Type of data	Method	Level of Analysis		Process	Outcome
		Individual	Team		
Qualitative	Protocol Analysis	X	X	X	
	Critical Incident Technique	X	X	X	
	Critical Decision Method	X	X	X	
	Concept Mapping	X			X
Quantitative	Behavioral Observation Scales (BOS)	X	X	X	
	Communication analysis		X	X	
	Event-based measurement		X	X	
	Pathfinder	X			X
	Self-report measures	X	X	X	X
	Automated Performance Recording	X	X	X	X

Table 1: Measurement methodologies with type of data, level of analysis, and phase of IPO model

<b>Measurement system requirement</b>	<b>Best Practice</b>
Multiple levels of measurement	<ol style="list-style-type: none"> <li>1. Performance measurement works best when it captures multiple dimensions of performance at appropriate levels of analysis.</li> <li>2. Performance measurement works best when event-based techniques are used to capture data at multiple levels of analysis.</li> <li>3. Performance measurement works best when multiple measures from various sources are captured.</li> <li>4. Performance measurement works best when a systematic plan is in place to integrate data from multiple measures.</li> </ol>
Address process as well as outcome	<ol style="list-style-type: none"> <li>5. Performance measurement works best when it captures the processes of performance.</li> <li>6. Performance measurement works best when expert models of the task are used to evaluate performance processes and outcomes.</li> <li>7. Performance measurement works best when the collection and transmission of objective measures is automated.</li> </ol>
Describe, evaluate, and diagnose performance	<ol style="list-style-type: none"> <li>8. Performance measurement works best when measures are descriptive of performance.</li> <li>9. Performance measurement works best when performance can be evaluated.</li> <li>10. Performance measurement works best when it is diagnostic—when it provides insight into the causes of performance.</li> <li>11. Performance measurement works best when it allows for performance evaluation to be partially or fully automated.</li> <li>12. Performance measurement works best when it allows for performance diagnosis to be partially or fully automated.</li> <li>13. Performance measurement works best when there is flexibility designed into embedded measures (different measures can be substituted).</li> <li>14. Performance measurement works best when it captures a broad spectrum of measures and the context of performance.</li> <li>15. Performance measurement works best when observers are trained to high levels of reliability.</li> <li>16. Performance measurement works best when observers use protocols.</li> </ol>

<p>Provide a basis for remediation</p>	<ol style="list-style-type: none"> <li>17. Performance measurement works best when it supports learning.</li> <li>18. Performance measurement works best when it allows for the automated and manual creation of AAR aids for training remediation and feedback.</li> <li>19. Performance measurement works best when it enables automated scaffolding and performance-based coaching.</li> <li>20. Performance measurement works best when it drives real-time corrective feedback.</li> <li>21. Performance measurement works best when it is integrated with trainer controls and feedback generation.</li> </ol>
--	--

Table 2. Best practices for performance measurement in training simulation.

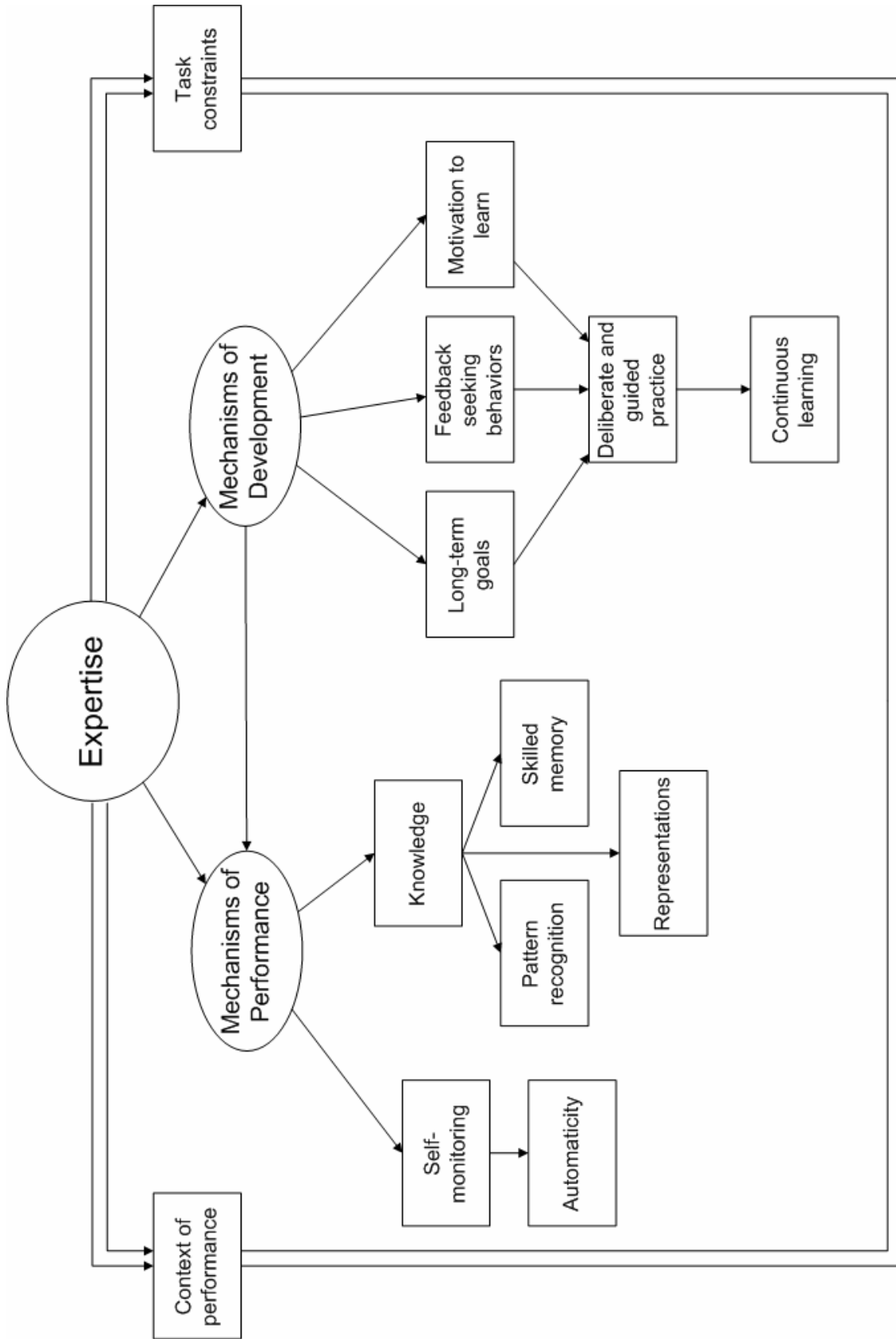


Figure 1. A framework of the components of expertise and its development (adapted from Salas & Rosen, in press)

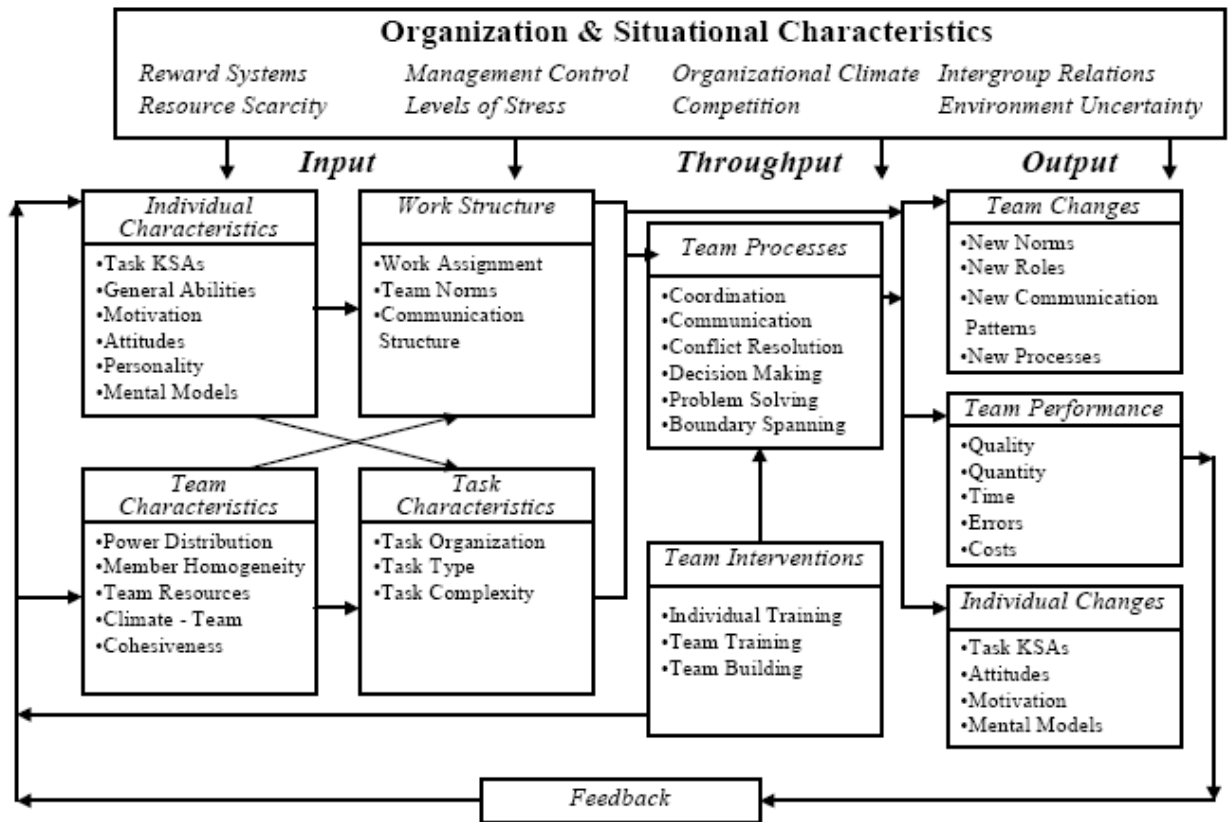


Figure 2: The team effectiveness framework (Tannenbaum, Beard, & Salas, 1992)

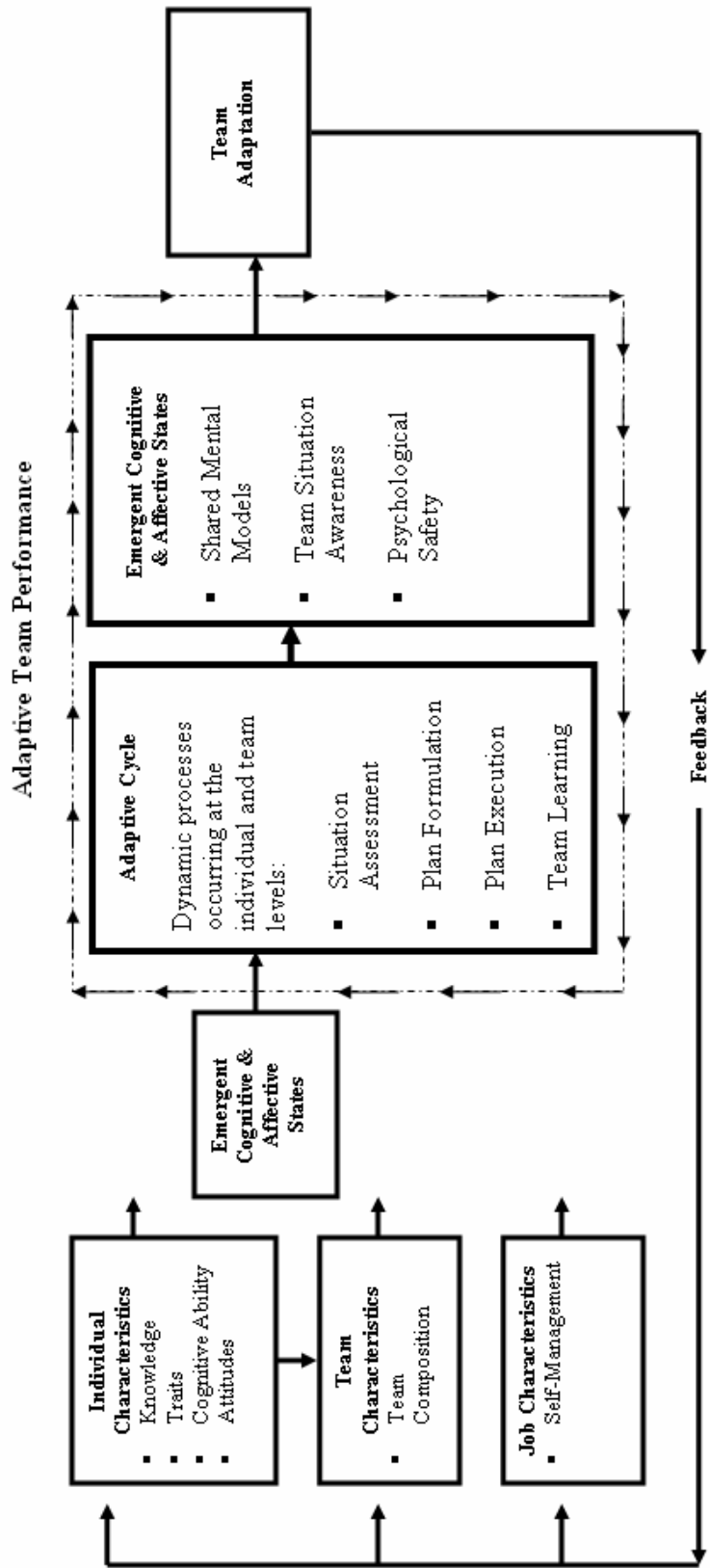


Figure 3: Team adaptation model (Burke, Stagl, Salas, Pierce, & Kendall, in press).

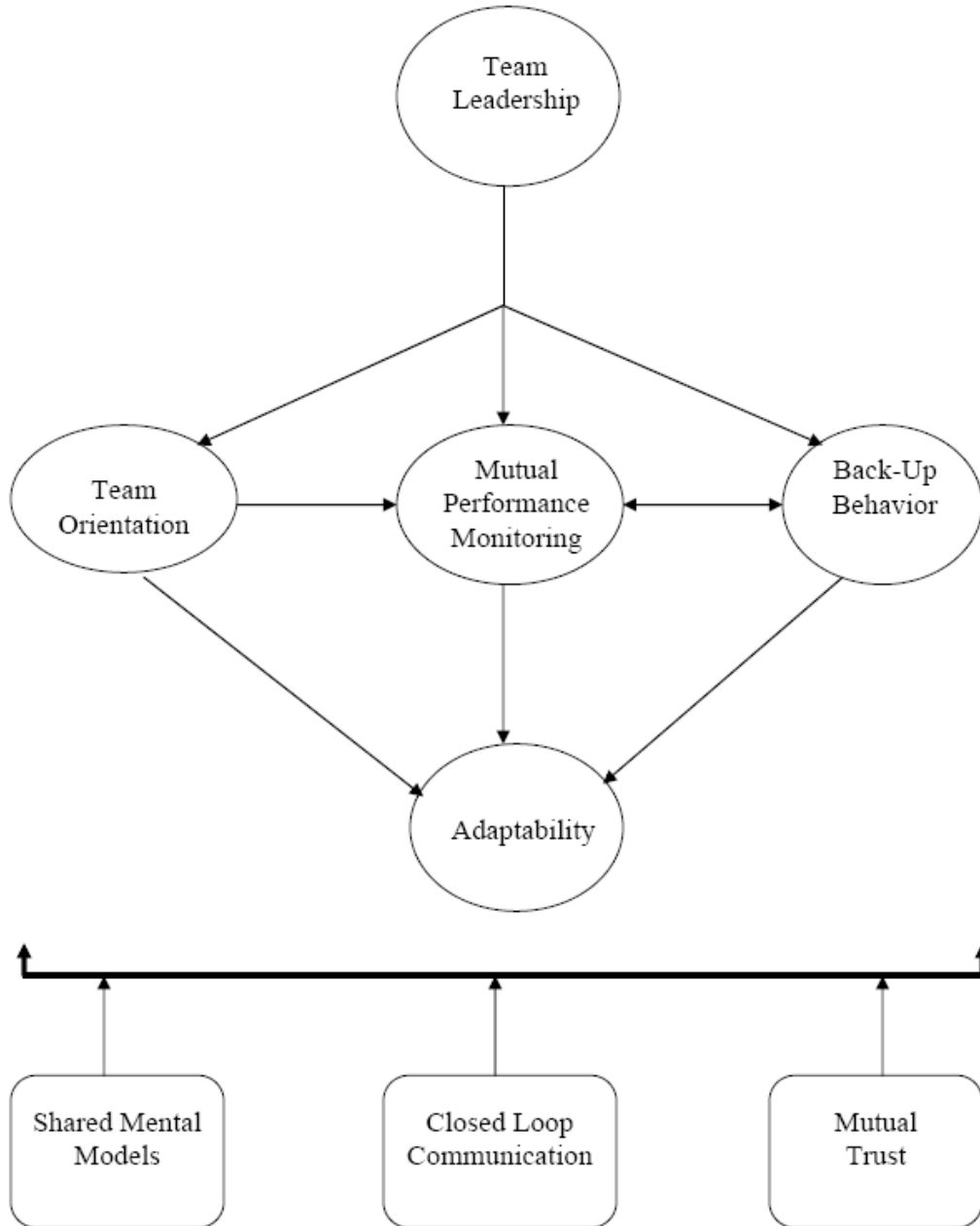


Figure 4: The big five teamwork model (Salas, Sims, & Burke, 2005)

### References

- Annett, J., Cunningham, D., & Mathias-Jones, P. (2000). A method for measuring team skills. *Ergonomics*, *43*, 1076-1094.
- Austin, J.R. (2003). Transactive memory in organizational groups: The effects of content, consensus, specialization, and accuracy on group performance. *Journal of Applied Psychology*, *88*(5), 866-878.
- Baddeley, A.D., & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval? *Memory and Cognition*, *21*, 146-155.
- Bainbridge, L., & Sanderson, P. (2005). Verbal protocol analysis. In J.R. Wilson and E.M. Corlett (Eds.), *Evaluation of Human Work: A practical ergonomics methodology* (3rd ed., pp. 159-184). Boca Raton, FL: Taylor & Francis.
- Bandow, D. (2001). Time to create sound teamwork. *The Journal for Quality and Participation*, *24*, 41-47.
- Banks, D., Bader, P., Fleming, P., Zaccaro, S., & Barber, H. (2001). *Leader adaptability: The role of work experiences and leader attributes*. Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.
- Baker, D.P., & Salas, E. (1992). Principles for measuring teamwork skills. *Human Factors*, *34*(4), 469-475.
- Bedard, J., & Chi, M. T. H. (1992). Expertise. *Current directions in psychological science*, *1*(4), 135-139.
- Berger, H. (1999). *Realtime feedback in truck driver training*. Paper presented at the Interservice/Industry Simulation, Training, and Education Conference.
- Bordage, G., & Zacks, R. (1984). The structure of medical knowledge in the memories of medical students and general practitioners: Categories and prototypes. *Medical education*, *18*, 406-416.
- Borman, W.C. (1991). Job behavior, performance, and effectiveness. In M. D. Dunette, and L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 271-326). Palo Alto, CA: Consulting Psychologists Press.
- Bowers, C.A., Jentsch, F., Salas, E., & Braun, C.C. (1998). Analyzing communication sequences for team training needs assessment. *Human Factors*, *40*(4), 672-679.
- Brannick, M.T., & Prince, C. (1997). Overview of team performance measurement. In M.T. Brannick, E. Salas, and C. Prince (Eds.), *Team Performance Assessment and Measurement: Theory, Methods, and Applications* (pp. 3-16). Mahwah, NJ: Erlbaum.
- Broadbent, D. (1972). *Decision and stress*. New York: Academic Press.
- Burke, C.S., Fiore, S.M., & Salas, E. (2004). The Role of Shared Cognition in Enabling Shared Leadership and Team Adaptability. In C.L. Pearce and J.A. Conger (Eds.), *Shared Leadership: Reframing the hows and whys of leadership* (pp. 103-121). Thousand Oaks, CA: Sage.
- Burke, C.S., Stagl, K.C., Salas, E., Pierce, L., & Kendall, D. (in press). Understanding team adaptation: A conceptual analysis & model. *Journal of Applied Psychology*.
- Cacciabue, P. C., & Hollnagel, E. (1995). Simulation of cognition: Applications. In J. M. Hoc, P.C. Cacciabue, and E. Hollnagel (Eds.), *Expertise and technology: Cognition and human-computer cooperation* (pp. 55-73). Mahwah, NJ: Erlbaum.

- Campbell, J.P. (1990). Modeling the performance prediction problem in Industrial and Organizational Psychology. In M.D. Dunette and L.M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology*. Palo Alto, CA: Consulting Psychologists Press.
- Campbell, J.P., McCloy, R.A., Oppler, S.H., & Sager, E. (1993). A theory of performance. In N. Schmitt and W. Borman (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco, CA: Jossey-Bass.
- Cannon-Bowers, J.A., Prince, C., Salas, E., Owens, J.M., Morgan, B.B., & Gonos, G.H. (1989). *Determining aircrew coordination training effectiveness*. Paper presented at the 11th annual meeting of the Interservice/Industry Training Systems Conference, Ft. Worth, TX.
- Cannon-Bowers, J.A., & Salas, E. (1997). A framework for developing team performance measures in training. In M.T. Brannick, E. Salas, and C. Prince (Eds.), *Team Performance and Measurement: Theory, Methods, and Applications* (pp. 45-62). Mahwah, NJ: Erlbaum.
- Cannon-Bowers, J.A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N.J.J. Castellan (Ed.), *Individual and group decision making* (pp. 221-246). Hillsdale, N.J.: Erlbaum.
- Card, S. K., Moran, T.P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Erlbaum.
- Carless, S.A., & De Paola, C. (2000). The measurement of cohesion in work teams. *Small Group Research*, 31(1), 71-88.
- Carley, K.M. (1997). Extracting Team Mental Models through Textual Analysis. *Journal of Organizational Behavior*, 18, 533-558.
- Carolan, T., MacMillan, J., Entin, E.B., Morley, R.M., Schreiber, B.T., Portrey, A., Denning, T., & Bennett, W. (2003). *Integrated performance measurement and assessment in distributed mission operations environments: Relating measures to competencies*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.
- Carron, A. V., & Brawley, L.R. (2000). Cohesion: Conceptual and measurement issues. *Small Group Research*, 31(1), 89-106.
- Chandler, T. N., & Owens, N.E. (2003). *Using performance based coaching techniques for intelligent tutoring systems*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.
- Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. J. Fletovich & R. R. Hoffman (Eds.), *The cambridge handbook of expertise and expert performance* (pp. 21-30). New York: Cambridge University Press.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science*, 5, 121-152.
- Cohen, S.G., & Bailey, D.E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, 23(3), 239-290.
- Cohen, M.S., Freeman, J.T., & Wolf, S. (1996). Metarecognition in time-stressed decision making: Recognizing, critiquing and correcting. *Human Factors*, 38(2), 206-219.

- Cooke, N.J., Salas, E., Cannon-Bowers, J.A., & Stout, R.J. (2000). Measuring team knowledge. *Human Factors*, 42, 151-173.
- Cooke, N.J., Salas, E., Kiekel, P.A., & Bell, B. (2004). Advances in measuring team cognition. In E. Salas and S.M. Fiore (Eds.), *Team Cognition: Understanding the Factors that Drive Process and Performance* (pp. 83-106). Washington, DC: APA.
- Cooke, N.J., & Schvaneveldt, R.W. (1988). Effects of computer programming experience on network representations of abstract programming requirements. *International Journal of Man-Machine Studies*, 29, 407-427.
- Cooke, N.J., Stout, R., & Salas, E. (1997). *Broadening the measurement of situational awareness through cognitive engineering methods*. Paper presented at the 41st Annual Meeting of the Human Factors and Ergonomics Society.
- Cooke, N.J., Stout, R.J., & Salas, E. (2001). A knowledge elicitation approach to the measurement of team situation awareness. In M. McNeese, E. Salas, and M.R. Endsley (Eds.), *New Trends in Cooperative Activities* (pp. 114-139). Santa Monica, CA: HFES.
- Cooper, R., & Sawaf, A. (1996). *Executive EQ: Emotional intelligence in leadership and organizations*. New York: Grosset/Putnam.
- Cooper, M., Viney, L., & McDermott, T. (2003). *Electronic Warfare Rangeless Embedded Training: A Cost Effective Training Approach*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.
- Cota, A.A., Evans, C.R., Dion, K.L., Kilik, L., & Longman, R.S. (1995). The structure of group cohesion. *Personality and Social Psychology Bulletin*, 21(6), 572-580.
- Drewes, P., & Gonzalez, A. (1997). *Automated performance monitoring*. Paper presented at the Interservice/Industry Simulation, Training and Education Conference, Orlando, FL.
- Driskell, J.E., & Salas, E. (1991). Group Decision Making Under Stress. *Journal of Applied Psychology*, 76(3), 473-478.
- Driskell, J.E., & Salas, E. (1992). Collective behavior and team performance. *Human Factors*, 34, 277-288.
- Driskell, J. E., & Salas, E. (Eds.). (1996). *Stress and human performance*. Mahwah, NJ: Erlbaum.
- Dwyer, D.J., Fowlkes, J.E., & Oser, R.L. (1997). Team performance measurement in distributed environments: The TARGETs methodology. In M.T. Brannick, E. Salas, and C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 137-153). Mahwah, NJ: Erlbaum.
- Dwyer, D.J., Oser, R.L., Salas, E., & Fowlkes, J.E. (1999). Performance measurement in distributed environments: Initial results and implications for training. *Military Psychology*, 11(2), 189-215.
- Dyer, J.L. (1984). Team Research and Team Training: a State of the Art Review. In F.A. Muckler (Ed.), *Human Factors Review* (pp. 285-323). Santa Monica: Human Factors Society.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32-64.

- Endsley, M. R. (1997). The role of situation awareness in naturalistic decision making. In C. E. Zsombok, and G. Klein (Eds.), *Naturalistic Decision Making*. Hillsdale, NJ: Erlbaum.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological review*, 102(2), 211-245.
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual review of psychology*, 47, 273-305.
- Ericsson, K.A., & Simon, H.A. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: The MIT Press.
- Ericsson, K. A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: an introduction. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 1-38). Cambridge, UK: Cambridge University Press.
- Evans, C.R., & Dion, K.L. (1991). Group cohesion and performance: a meta-analysis. *Small Group Research*, 22(2), 175-186.
- Faraj, S.A., & Sproull, L. (2000). Coordinating expertise in software development teams. *Management Science*, 46, 1554-1568.
- Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 41-67). New York: Cambridge University Press.
- Fiore, S. M., & Schooler, J.W. (2004). Process mapping and shared cognition: Teamwork and the development of shared problem models. In E. Salas, and S.M. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 133-152). Washington, D.C.: American Psychological Association.
- Fivar, G. (1980). *The critical incident technique: A bibliography* (2nd ed.). Palo Alto, CA: American Institute for Research.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-development inquiry. *American Psychologist*, 34(10), 906-911.
- Fleishman, E.A., Mumford, M.D., Zaccaro, S.J., Levin, K.Y., Korotkin, A.L., & Hein, M.B. (1991). Taxonomic efforts in the description of leader behavior: A synthesis and functional interpretation. *Leadership Quarterly*, 4, 245-287.
- Fleming, B., & Barto, J., & Johnson, C. (1998). *Joint collective performance measures: Performance validation through measurement*. Paper presented at the Interservice/Industry Training, Simulation and Education Conference, Orlando, FL.
- Foushee, H.C., & Manos, K.L. (1981). *Information transfer within the cockpit: Problems in intra cockpit communications* (No. NASA TP-1875). Moffett Field, CA: NASA-Ames Research Center.
- Fowlkes, J. E., Dwyer, D.J., Milham, L.M., Burns, J.J., & Pierce, L.G. (1999). *Team skills assessment: A test and evaluation component for emerging weapon systems*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference.

- Fowlkes, J. E., Dwyer, D. J., Oser, R. L., & Salas, E. (1998). Event-based approach to training (EBAT). *The international journal of aviation psychology*, 8(3), 209-221.
- Fowlkes, J.E., Lane, N.E., Salas, E., Franz, T., & Oser, R. (1994). Improving the measurement of team performance: The TARGETs methodology. *Military Psychology*, 6, 47-61.
- Frank, G., Meeds, H., Perkins, K., Whiteford, B., Arnold, P., Hubal, R., Sonker, P., Presley, T., & Jones, R. (2004). *Performance assessment for distributed learning using after action review reports generated by simulations*. Paper presented at the Interservice/Industry Training, Simulation and Education Conference, Orlando, FL.
- Gersick, C.J. (1988). Time and Transition in Work Teams: Toward a New Model of Group Development. *Academy of Management Journal*, 31(1), 9-41.
- Glaser, R. (1987). Thoughts on expertise. In C. Schooler & K. W. Schaie (Eds.), *Cognitive functioning and social structure over the life course* (pp. 81-94). Norwood, NJ: Ablex.
- Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser & M. Farr (Eds.), *The nature of expertise* (pp. xv-xxviii). Hillsdale, NJ: Erlbaum.
- Goldsmith, T.E., Johnson, P.J., & Acton, W.H. (1991). Assessing structural knowledge. *Journal of Educational Psychology*, 83, 88-97.
- Goldsmith, T.E., & Kraiger, K. (1996). Applications of structural assessment to training evaluation. In J. K. Ford, S.W.J. Kozlowski, and E. Salas (Eds.), *Improving training effectiveness in organizations* (pp. 73-96). Hillsdale, NJ: Erlbaum.
- Graham, J., Schneider, M., Bauer, A., Bessiere, K. & Gonzalez, C. (2004). *Shared mental models in military command and control organizations: Effect of social network distance*. Paper presented at the Human Factors and Ergonomics Society 48th Annual Meeting, New Orleans, LA.
- Gray, B., Bougon, M.G., & Donnellon, A. (1985). Organizations as constructions and destructions of meaning. *Journal of Management*, 11, 83-98.
- Grigorenko, E. L., & Sternberg, R.J. (1998). Dynamic testing. *Psychological Bulletin*, 124, 75-111.
- Gully, S. M., Devine, D.J., & Whiney, D.J. (1995). A meta-analysis of the cohesion and performance: Effects of level of analysis and task interdependence. *Small Group Research*, 26, 497-520.
- Hackman, J.R. (1983). *A Normative Model of Work Team Effectiveness* (No. 2). New Haven, CT: Yale University.
- Hackman, J.R. (2002). *Leading teams: setting the stage for great performances*. Boston: Harvard Business School Press.
- Hancock, P. A., & Meshkati, N. (Eds.). (1988). *Human Mental Workload*. Amsterdam: North-Holland.
- Hancock, P. A., & Warm, J. (1989). A Dynamic Model of Stress and Sustained Attention. *Human Factors*, 31(5), 519-537.
- Hobson, C.J., & Gibson, F.W. (1984, March). Capturing supervisor rating policies: A way to improve performance appraisal effectiveness. *Personnel Administrator*, 29(3), 59-68.
- Hockey, R., & Hamilton, P. (1983). The Cognitive Patterning of Stress States. In G. R. J. Hockey (Ed.), *Stress and Fatigue in Human Performance*: Wiley.

- Hoffman, R.R., Crandall, B., & Shadbolt, N. (1998). Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors*, 40(2), 254-276.
- Hoffman, R. R., Feltovich, P. J., & Ford, K. M. (1997). A general conceptual framework for conceiving of expertise and expert systems. In P. J. Feltovich, K. M. Ford & R. R. Hoffman (Eds.), *Expertise in context* (pp. 543-580). Menlo Park, CA: AAAI Press/MIT Press.
- Jentsch, F., Sellin-Wolters, S., Bowers, C.A., & Salas, E. (1995). *Crew coordination behaviors as predictors of problem detection and decision making times*. Paper presented at the Human Factors and Ergonomics Society 30th Annual Meeting, Santa Monica, CA.
- Johnson-Laird, P.N. (1983). *Mental Models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Johnston, J.H., Cannon-Bowers, J.A., & Smith-Jentsch, K.A. (1995). *Event-based performance measurement system for shipboard command teams*. Paper presented at the First International Symposium on Command and Control Research and Technology, Washington, DC.
- Johnston, J.H., Smith-Jentsch, K.A., Cannon-Bowers, J.A. (1997). Performance measurement tools for enhancing team decision-making training. In M.T. Brannick, E. Salas, and C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 311-327). Mahwah, NJ: Erlbaum.
- Jones, G., & George, J. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review*, 23, 531-546.
- Kane, J. S., & Woehr, D. J. (2006). Performance measurement reconsidered: An examination of frequency estimation as a basis for assessment. In W. J. Bennett (Ed.), *Performance measurement: Current perspectives and future challenges* (pp. 77-110). Mahwah, NJ: Erlbaum.
- Kanki, B.G., Lozito, S., & Foushee, H.C. (1989). Communication indices of crew coordination. *Aviation, Space, & Environmental Medicine*, 60, 56-60.
- Keikel, P.A., Cooke, N.J., Foltz, P.W., & Shope, S.M. (2001). Automating measurement of team cognition through analysis of communication data. In M. J. Smith, G. Salvendy, D. Harris, and R.J. Koubek (Eds.), *Usability Evaluation and Interface Design*. Mahwah, NJ: Erlbaum.
- Kelly, M. (1999). *After action review in synthetic environment based training systems: A training system not a technology*. Paper presented at the Interservice/Industry Simulation, Training, and Education Conference.
- Klein, G. (1993). A recognition primed decision (RPD) model of rapid decision making. In G. Klein, J. Orasanu, R. Calderwood, and C.E. Zsombok (Eds.), *Decision making in action* (pp. 138-147). Norwood, NJ: Ablex.
- Klein, G.A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(3), 462-472.
- Klein, K., & Kozlowski, S.W.J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein (Ed.),

- Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions* (pp. 3-90). San Francisco, CA: Jossey-Bass.
- Klein, G., Ross, K.G., Moon, B.M., Klein, D.E., Hoffman, R.R., & Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems*, 18(3), 81-85.
- Kleinman, D. L., & Serfaty, D. (1989). *Team Performance Assessment in Distributed Decision-making*. Paper presented at the Symposium on Interactive Networked Simulation for Training, Orlando.
- Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, 20(2), 403-437.
- Kluwe, R.H. (1987). Executive decisions and regulation of problem-solving behavior. In F.E. Weinert and R.H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 31-64). Hillsdale, NJ: Erlbaum.
- Kozlowski, S.W.J. (1998). Training and developing adaptive teams: Theory, principles, and research. In J.A. Cannon-Bowers and E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 115-153). Washington, D.C.: APA.
- Kozlowski, S.W.J., & Bell, B.S. (2003). Work groups and teams in organizations. In W. C. Borman, D.R. Ilgen, and R.J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 333-375). London: Wiley.
- Kozlowski, S.W.J., Gully, S.M., Nason, E.R., & Smith, E.M. (1999). Developing adaptive teams: A theory of compilation and performance across levels and time. In D. R. Ilgen, and E.D. Pulakos (Eds.), *The Changing Nature of Work and Performance: Implications for Staffing, Personnel Actions, and Development*. San Francisco, CA: Jossey-Bass.
- Kozlowski, S.W.J., Gully, S.M., & Salas, E. (1996). Team Leadership and Development: Theory, principles, and guidelines for training leaders and teams. In M.M. Beyerlein and A. Dougals (Eds.), *Advances in Interdisciplinary Studies of Work Teams: Team Leadership* (Vol. 3, pp. 253-291). Greenwich, CT: JAI Press.
- Kozlowski, S.W.J., Gully, S.M., Smith, E.M., Nason, E.R., & Brown, K.G. (1995). *Sequenced mastery training and advanced organizers: Effects on learning, self-efficacy, performance, and generalization*. Paper presented at the Symposium conducted at the Tenth Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Kozlowski, S.W.J., & Klein, K.J. (2000). A Multilevel Approach to Theory and Research in Organizations: Contextual, Temporal, and Emergent Processes. In K.J. Klein and S.W.J. Kozlowski (Eds.), *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions* (pp. 3-90). San Francisco, CA: Jossey-Bass.
- Leddo, J., Zhang, Z., & Pokorny, R. (1998). *Automated Performance Assessment Tools*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.
- Lesgold, A. M., Rubinson, H., Feltovich, P. J., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. T. H. Chi, R. Glaser & M. Farr (Eds.), *The nature of expertise* (pp. 311-342). Hillsdale, NJ: Erlbaum.

- Lewis, K. (2000). *Transactive memory and performance of management consulting teams: Examining construct and predictive validity of a new scale*. Paper presented at the Annual meeting of the Academy of Management, Toronto, Canada.
- Lott, A.J., & Lott, B.D. (1965). Group cohesiveness as interpersonal attraction: A review of relationships with antecedent and consequent variables. *Psychological Bulletin*, *64*, 259-306.
- MacBryde, J., & Mendibil, K. (2003). Designing performance measurement systems for teams: theory and practice. *Management Decision*, *41*(8), 722-733.
- Marks, M. A. (2000). A critical analysis of computer simulations for conducting team research. *Small Group Research*, *31*, 653-657.
- Marks, M.A., Zaccaro, S.J., & Mathieu, J.E. (2000). Performance implications of leader briefings and team interaction training for team adaptation to novel environments. *Journal of Applied Psychology*, *85*, 971-986.
- Marks, M.A., Mathieu, J.E., & Zaccaro, S.J. (2001). A Temporally Based Framework and Taxonomy of Team Processes. *Academy of Management Review*, *26*, 356-376.
- Marsh, S.M., Kiechel, L., Boyce, L.A., & Zaccaro, S.J. (2001). *Leadership and Team Adaptation: Examining the Attributes and Training Strategies that Promote Effective Performance in Dynamic Environments*. Paper presented at the 13th Annual Conference of Society for Industrial and Organizational Psychology, San Diego, CA.
- Marhsall-Mies, J.C., Fleishman, E.A., Martin, J.A., Zaccaro, S.J., Baughman, W.A., & McGee, M.L. (2000). Development and evaluation of cognitive and metacognitive measures for predicting leadership potential. *Leadership Quarterly*, *11*, 135-154.
- Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Salas, E., & Cannon-Bowers, J. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*(2), 273-283.
- Mazzoni, G., & Nelson, T.O. (Eds.). (1998). *Metacognition and cognitive neuropsychology: Monitoring and control processes*. Mahwah, NJ: Lawrence Erlbaum Associates.
- McCloy, R. A., Campbell, J.P., & Cudeck, R. (1994). A Confirmatory Test of a Model of Performance Determinants. *Journal of Applied Psychology*, *79*(4), 493-505.
- Meister, D. (1985). *Behavioral Analysis and Measurement Methods*. New York: Wiley.
- Meliza, L. L., & Brown, B. (1996). *Increasing speed and flexibility of feedback systems for DIS Exercises*. Paper presented at the Interservice/Industry Training, Simulation and Education Conference.
- Meliza, L. L., & Paz, B. (1996). *Integrating Exercise Control and Feedback Systems in DIS*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference.
- Mohammed, S., & Dumville, B.C. (2001). Team mental models in a team knowledge framework: Expanding theory and measure across disciplinary boundaries. *Journal of Organizational Behavior*, *22*(2), 89-103.
- Mohammed, S., Klimoski, R., & Rentsch, J.R. (2000). The measurement of team mental models: We have no shared schema. *Organizational Research Methods*, *3*(123-165).

- Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological bulletin*, *132*(2), 297-326.
- Moreland, R.L. (1999). Transactive memory: Learning who knows what in work groups and organizations. In L.L. Thompson and J.M. Levine (Eds.), *Shared cognition in organizations: The management of knowledge* (pp. 3-31). Mahwah, NJ: Lawrence Erlbaum Associates.
- Moreland, R.L., Argote, L., & Krishnan, R. (1996). Socially shared cognition at work: Transactive memory and group performance. In J.L. Nye and A.M. Brower (Eds.), *What's social about social cognition? Research on socially shared cognition in small groups* (pp. 57-84). Thousand Oaks, CA: Sage Publications.
- Morgan, B.B., Jr., Glickman, A.S., Woodward, E.A., Blaiwes, A.S., & Salas, E. (1986). *Measurement of Team Behaviors in a Navy Environment* (No. 86-014). Orlando, FL: Naval Training Systems Center.
- Morgan, B.B., Jr., Lassiter, D.L., & Salas, E. (1989). *Networked simulation application for team training and performance research*. Paper presented at the meeting for Interactive Networked Simulation in Training, Orlando, FL.
- Morgan, B.B., Jr., Salas, E., & Glickman, A.S. (2001). An Analysis of Team Evolution and Maturation. *The Journal of General Psychology*, *120*(3), 277-291.
- Motowildo, S.J. (2003). Job performance. In W.C. Borman, D.R. Ilgen, and R.J. Klimoski (Eds.), *Comprehensive Handbook of Psychology* (Vol. 12: Industrial and organizational psychology, pp. 39-53). New York: Wiley.
- Murphy, K. R., & Constans, J.L. (1987). Behavioral anchors as a source of bias in rating. *Journal of Applied Psychology*, *72*, 573-577.
- Murphy, K. R., & Pardaffy, V.A. (1989). Bias in behaviorally anchored rating scales: Global or scale-specific? *Journal of Applied Psychology*, *74*, 343-346.
- Orasanu, J. (1990). *Shared mental models and crew performance*: Laboratory of Cognitive Science, Princeton University, NJ.
- Orasanu, J., & Fischer, U. (1991). *Information transfer and shared mental models of decision making*. Paper presented at the Sixth International Symposium on Aviation Psychology, Columbus, OH.
- Oser, R. L., Cannon-Bowers, J. A., Salas, E., & Dwyer, D. J. (1999). Enhancing human performance in technology-rich environments: Guidelines for scenario-based training. In *Human/technology interaction in complex systems* (Vol. 9, pp. 175-202): JAI Press.
- Pearce, C.L., & Sims, H.P. (2002). Vertical versus shared leadership as predictors of the effectiveness of change management teams: An examination of aversive, directive, transactional, transformational and empowering leader behaviors. *Group Dynamics: Theory, Research, and Practice*, *6*(2), 172-197.
- Piotrowski, M.J., Barnes-Farrell, J.L., & Esrig, F.H. (1989). Behaviorally anchored bias: A replication and extension of Murphy and Constans. *Journal of Applied Psychology*, *74*, 823-826.
- Porter, C.O., Hollenbeck, J.R., Ilgen, D.R., Ellis, A.P., West, B.J., & Moon, H. (2003). Backing up behaviors in teams: the role of personality and legitimacy of need. *Journal of Applied Psychology*, *88*(3), 391-403.
- Posner, M. I. (1978). *Chronometric explorations of the mind*. Hillsdale, NJ: Erlbaum.

- Prince, C., & Salas, E. (1997). Situation assessment for routine and flight decision making. *International Journal of Cognitive Ergonomics*, *1*, 315-204.
- Redden, E.S., Sheehy, J.B., & Bjorkman, E.A. (2004). The study and measurement of human performance by military service laboratories. In J. M. Ness, V. Tepe, and D. Ritzer (Eds.), *Advances in human performance and cognitive engineering research* (Vol. 5, pp. 517-579). US: Elsevier Science/JAI Press.
- Salas, E., Burke, C.S. & Stagl, K.C. (2004). Developing teams and team leaders: Strategies and principles. In D. Day, S.J.W. Zaccaro, and S.M. Halpin (Eds.), *Leader development for transforming organizations: Growing Leaders for Tomorrow* (pp. 325-355). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Salas, E., & Cannon-Bowers, J.A. (2001). The science of team training: A decade of progress. *Annual Review of Psychology*, *52*(471-499).
- Salas, E., Dickinson, T., Converse, S., & Tannenbaum, S. (1992). Toward and Understanding of Team Performance and Training. In R. Swezey and E. Salas (Eds.), *Teams: Their training and performance*. Norwood, N.J.: Ablex Publishing.
- Salas, E., & Fiore, S.M. (2004). Why team cognition? An overview. In E. Salas and S.M. Fiore (Eds.), *Team cognition* (pp. 3-8). Washington, D.C.: American Psychological Association.
- Salas, E., Priest, H.A., & Burke, C.S. (2005). Teamwork and Team Performance Measurement. In J.R. Wilson and N. Corlett (Eds.), *Evaluation of Human Work* (3rd ed., pp. 793-808). Boca Raton, FL: Taylor & Francis.
- Salas, E., Prince, C., Baker, D.P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human Factors*, *37*(1), 123-136.
- Salas, E., & Rosen, M.A. (in press). Experts at work: Principles developing expertise in organizations. In S. W. J. Kozlowski, and Salas, E. (Eds.), *Learning, training, and development in organizations*. Mahwah, NJ: Erlbaum.
- Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D., & Howse, W. R. (2007, in press). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation, Space, and Environmental Medicine Special Supplement on Operational Applications of Cognitive Performance Enhancement Technologies*, *78*(5), xxxx.
- Salas, E., Stagl, K.C., & Burke, C.S. (2004). 25 Years of Team Effectiveness in Organizations: Research Themes and Emerging Needs. *International Review of Industrial and Organizational Psychology*, *19*, 47-91.
- Sanderson, P.M., & Fisher, C. (1994). Exploratory Sequential Data Analysis: Foundations. *Human-Computer Interaction*, *9*, 251-317.
- Schvaneveldt, R.W. (1990). Proximities, networks, and schemata. In R.W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 135-148). Norwood, NJ: Ablex.
- Schraagen, J.M., Chipman, S.F., & Shalin, V.J. (Eds.). (2000). *Cognitive Task Analysis*. Mahwah, NJ: Erlbaum.
- Schreiber, B.T., Watz, E.A., & Winston, B. (2003). *Objective human performance measurement in a distributed environment: Tomorrow's needs*. Paper presented at the Interservice/Industry Training, Simulation and Education Conference, Orlando, FL.

- Shadbolt, N. (2005). Eliciting expertise. In J.R. Wilson and N. Corlett (Eds.), *Evaluation of Human Work* (pp. 185-218). Boca Raton, FL: Taylor & Francis.
- Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American scientist*, *61*, 393-403.
- Smith, P.C., & Kendall, L.M. (1963). Retranslations or expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Scales*, *47*, 149-155.
- Smith-Jentsch, K.A., Tannenbaum, S.I., & Cannon-Bowers, J.A. (1995). *Training team leaders briefing skills. The science and practice of enhancing teamwork in organizations*. Paper presented at the Symposium conducted at the meeting of the American Psychological Association, Toronto, Canada.
- Sonnentag, S., Niessen, C., & Volmer, J. (2006). Expertise in software design. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 373-387). Cambridge, UK: Cambridge University Press.
- Stacy, W., Freeman, J., Lackey, S., & Merket, D. (2004). *Enhancing Simulation-Based Training with Performance Measurement Objects*. Paper presented at the Interservice/Industry Simulation, Training, and Education Conference, Orlando, FL.
- Sternberg, R.J. (1998). Metacognition, abilities, and developing expertise: What makes an expert student? *Instructional Science*, *26*, 127-140.
- Sternberg, R. J. (1997). Cognitive conceptions of expertise. In P. J. Fletovich, K. M. Ford & R. R. Hoffman (Eds.), *Expertise in context* (pp. 149-162). Menlo Park, CA: AAI Press/MIT Press.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, *30*, 276-315.
- Stout, R.J., Cannon-Bowers, J.A., & Salas, E. (1996). The role of shared mental models in developing team situational awareness: Implications for training. *Training Research Journal*, *2*, 85-116.
- Swan, J.A. (1995). Exploring knowledge and cognitions in decisions about technological innovations--Mapping managerial cognitions. *Human Relations*, *48*, 1241-1270.
- Tagger, S., & Seigts, G.H. (2003). Leader and staff role efficacy as antecedents of collective-efficacy and team performance. *Human Performance*, *16*, 131-156.
- 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.
- Tesluk, P.E., Mathieu, J.E., Zaccaro, S.J., & Marks, M. (1997). Task and aggregation issues in the analysis and assessment of team performance. In M. T. Brannick, E. Salas, and C. Prince (Eds.), *Team Performance Assessment and Measurement* (pp. 197-224). Mahwah, NJ: Erlbaum.
- Tubré, T., Arthur, W. Jr., & Bennett, W. Jr. (2006). General models of job performance: Theory and practice. In *Performance measurement: Current perspectives and future challenges* (pp. 175-203). Mahwah, NJ: Erlbaum.
- Webber, S. S. (2002). Leadership and trust facilitating cross-functional team success. *Journal of Management Development*, *21*, 201-214.

- Wegner, D.M. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen and G.R. Goethals (Eds.), *Theories of group behavior*. New York, NY: Springer-Verlag.
- Weick, K.E., & Bougon, M.G. (1986). Organizations as cognitive maps: Charting ways to success and failure. In H.P. Sims and D.A. Gioia (Eds.), *The thinking organization: Dynamics of organizational social cognition* (pp. 102-135). San Francisco, CA: Jossey Bass.
- Wickens, C. D. (1984). Processing Resources in Attention. In R. Parasuraman, and R. Davies (Eds.), *Varieties of Attention*. New York, NY: Academic Press.
- Wickens, C. D. (1987). Information Processing, Decision-Making, and Cognition. In G. Salvendy (Ed.), *Handbook of Human Factors* (pp. 72-107). New York: Wiley.
- Wickens, C. D., & Carswell, C.M. (1997). Information processing. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 89-129). New York: John Wiley & Sons, Inc.
- Wickens, C. D., & Hollands, J.G. (2000). *Engineering Psychology and Human Performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Yerkes, R. M., & Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurological Psychology*, 18, 459-482.
- Yoo, Y., & Kanawattanachai, P. (2001). Developments of transactive memory systems and collective mind in virtual teams. *International Journal of Organizational Analysis*, 9(2), 187-208.
- Zaccaro, S.J., Rittman, A.L., & Marks, M.A. (2001). Team leadership. *Leadership Quarterly*, 12, 451-483.
- Zachary, W., Bilazarian, P., Burns, J., & Cannon-Bowers, J.A. (1997). *Advanced embedded training concepts for shipboard systems*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.
- Zeitz, C. M. (1997). Some concrete advantages of abstraction: How experts' representations facilitate reasoning. In P. J. Feltovich, K. M. Ford & R. R. Hoffman (Eds.), *Expertise in context* (pp. 43-65). Menlo Park, CA: AAAI Press/MIT Press.