# A KNOWLEDGE ELICITATION APPROACH TO THE MEASUREMENT OF TEAM SITUATION AWARENESS


**NANCY J. COOKE**

New Mexico State University, Las Cruces, NM, USA

**RENE'E STOUT**

AlignMark, Maitland, FL, USA

and

**EDUARDO SALAS**

University of Central Florida, Orlando, FL, USA

Contact: Nancy J. Cooke, Ph.D.
Department of Psychology
New Mexico State University
Box 30001/3452
Las Cruces, NM 88003
Voice: 505-646-1630
Fax: 505-646-6212
Email: cooke@crl.nmsu.edu


Running Title: Measurement of Team SA

## INTRODUCTION

Recent research in human factors and particularly in aviation applications has relied heavily on the concept of SA, or situation awareness. Endsley (1988) defines SA as "the perception 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" (p. 97). Deficiencies in SA have been linked to a number of aviation accidents (Hartel, Smith, & Prince, 1991) and investigators in domains outside of aviation have also embraced this construct as one that is critical for effective operator performance in other environments (e.g., Gaba, Howard, & Small, 1995). Consequently, variations in operator performance have often been explained in terms of differences in operators' SA (e.g., Prince, Chidester, Bowers, & Cannon-Bowers, 1992) and design decisions have been made in order to improve SA (e.g., Endsley, 1988). Recently, investigators working in the team arena have proposed that teams can have SA that is something more than the sum of the SA of each individual team member (Prince, Stout, & Salas, 1994; Salas, Prince, Baker, & Shrestha, 1995).

Salas, Dickinson, Converse, and Tannenbaum (1992) define a team as "a distinguishable set of two or more people who interact dynamically, interdependently, and adaptively toward a common and valued goal/object/mission, who have each been assigned specific roles or functions to perform, and who have a limited life span of membership" (p. 126-127). Thus, team SA involves the team's assessment (i.e., perception, comprehension, and projection) of the current situation, which can include the surrounding environment (including any equipment or systems), the task, and the team itself.

In this paper we approach the measurement of team SA through the application of methods traditionally used for knowledge elicitation. Among other contributions, these methods allow for elicitation as well as assessment, and focus on underlying knowledge and cognition rather than outcome performance. It is our position that this elicitation approach not only provides measures of team SA, but that it also overcomes some limitations associated with the current measurement of individual SA as a result of current emphases on assessment and outcome performance. Specifically, the elicitation of team SA can augment the current assessment methods in order to provide information necessary for applications beyond system evaluation such as training and detailed design decisions and for more thoroughly understanding the cognitive underpinnings of team SA. More broadly, this approach can be applied to the measurement of team cognition.

## A FRAMEWORK FOR MEASURING TEAM SA

The selection, application, and adaptation of methods for the measurement of team SA requires guidance from existing conceptualizations, theories, and models of individual SA and team SA. There have been many very good theoretical treatments of these concepts (e.g., Endsley, 1995a; Salas et al., 1995; Stout, Cannon-Bowers, & Salas, 1996), as well as papers summarizing and integrating the varied definitions of SA (e.g., Durso & Gronlund, 1999; Sarter & Woods, 1991; Shrestha, Prince, Baker, & Salas, 1995). Rather than belabor specific nuances or variations of these positions, we believe it is more fruitful to forge ahead with the development and evaluation of measures that will then aid in refining conceptualizations of team SA. Therefore, in this section we propose a preliminary framework for the purpose of identifying starting points in the measurement of team SA. This framework relies heavily on the theoretical work of Salas, et al. (1995), Stout, et al., (1996), and Endsley (1995a) as points of departure. As this framework is described, we list aspects of team SA (i.e., precursors or products) that we target in the course of its measurement.

The phenomenon of team SA is multifaceted in that it involves the integration of a number of cognitive structures and processes as represented in Figure 1. Individual situation models and shared mental models can be considered precursor products of team SA in that they are cognitive structures that are brought to the task of interpreting the situation by team members.

[Insert Figures 1 and 2 about here]

Individual *situation models* are the products of individual SA for each team member (see Figure 2). Each situation model can be considered a snapshot of the individual's understanding of the situation at a single point in time (Endsley, 1995a) and can thus, affect the individual's performance. The process involved in deriving this understanding or situation model is called situation assessment in which the individual's knowledge, often referred to as a *mental model*, is integrated with cues or information in the environment in order to understand, explain, and predict in this context (Endsley, 1995a; Salas, et al., 1995; Stout, et al., 1996). The current understanding of the situation (i.e., the situation model) can alter the knowledge associated with the mental model. Further, feedback from performance can impact the situation assessment process directly and via its impact on the mental model. In addition, performance outcomes themselves serve as additional environmental cues for further situation assessment. Situation assessment is very much like the cue/pattern assessment process that experts use to quickly understand and act on the any given situation (e.g., Chase & Simon, 1973a; 1973b). In terms of assessing a team's SA, the factor most directly relevant at the individual level is the situation model. However, if one was interested in training or impacting team SA via the individual situation models, other aspects of individual SA (individual mental models, cues, and the situation assessment process) would likewise be critical and so, although only indirectly relevant to team SA, are included in Table 1 as targets of team SA measurement.

[Insert Table 1 about here]

The *shared mental model* is another precursor to team SA (see Figure 1). Not only do team members each bring to the task momentary or transient understandings of the current situation (i.e., individual situation models), but they also bring relevant individual knowledge pertinent to the environment, task, or team members that is long-term in nature (i.e., individual mental models). Thus, the framework proposed here distinguishes at both the individual and team levels the momentary *understanding* of the situation which is captured by situation models and the longer-term *knowledge* brought to bear on this understanding represented in mental models.

Further, what is thought to be critical to team SA is more than the aggregation of the team's individual mental models, but more of a holistic representation of this knowledge called a shared mental model (Blickensderfer, Cannon-Bowers, & Salas, 1997a; Cannon-Bowers, Salas, & Converse, 1993; Kraiger, Krause, & Hart, 1996; Kraiger & Wenzel, 1997; Rentsch & Hall, 1994; Rouse, Cannon-Bowers, & Salas, 1992; Stout et al., 1996). This holistic representation emphasizes the importance of integration of knowledge across team members and issues such as knowledge similarity and distribution. According to Stout et al. (1996), shared mental models can have content relevant to the overall task, team goals, team member roles, individual tasks, team members themselves, and information members need to complete a task and when it is needed. This knowledge can be declarative, procedural, or strategic in nature. One implication of shared mental models is that training interventions that extend beyond the background knowledge of individual team members, but instead target how that knowledge is distributed across team members, should be effective in modifying the team's performance associated with team SA. Further, one can distinguish the existence of knowledge at the team level from the subset of that knowledge that is available to impact team action (influenced by environmental cues and team behaviors), the latter of which is assumed in the holistic knowledge representation in Figure 1.

Whereas individuals integrate knowledge and environmental cues using the process of situation assessment, teams integrate the individual situation models and shared mental model via *team process behaviors*. These include behaviors such as communication, coordination, planning, and leadership (Shrestha et al., 1995; Stout, 1995). Prince, et al. (1994, p. 20-23) list specific team behaviors associated with SA. These behaviors serve to explicitly or implicitly integrate the information and knowledge of individual team members. Communication, for example, is a team process behavior that affects and is affected by team SA (Salas, et al., 1995). Communication enables team members to share information from their individual situation models so that it will become part of the team's situation model. In turn, changes in the team's situation model can influence the individual understanding, which in turn affects what is communicated and how it is communicated. Information can be given or received either verbally or nonverbally. According to Shrestha, et al., (1995) teams can also coordinate implicitly which is revealed when the actions of a team member are anticipated. Together, the shared mental model and the team process behaviors are aspects of team SA that go beyond the mere aggregation of individual SA and that are thus, potential targets for training interventions at the team level. Note that although, environmental cues play a role in the construction of individual situation models and thus indirectly affect process behaviors, they are also likely to impact the team process behaviors more directly, though this factor is not represented in Figure 1.

The product that results from the integration of situation models and the shared mental model via team process behaviors is the *team situation model* which represents the momentary understanding of the current situation (i.e., environment, task, team) by the team as a whole. This is the end product of team SA. Like the shared mental model, this shared understanding does not reside within a team member, but represents the result of integrating information and knowledge across team members. Also, like the shared mental model, the team situation model is thought to be distinct from the aggregation of individual situation models (see Hutchins (1991) for evidence supporting this assumption). For instance, the co-pilot may be aware of an impending threat to the aircraft, whereas the pilot is unaware of this threat. As a team, the pilot and co-pilot need to take action and the action will reflect the team situation model or how the two conflicting pieces of information are integrated to come to a team understanding of the situation. That is, despite the conflicting individual situation models, good team process behaviors such as communication and coordination can result in a unified and accurate team situation model. This team model should, in turn, influence one or both of the individual models and perhaps even the team process behaviors. Team understanding that is shared, or at least compatible, enables a team to coordinate quickly and implicitly when the environment restricts overt or more explicit strategizing (Cannon-Bowers, et al., 1993).

Note that there are multiple possibilities for feedback in this framework. First, as alluded to in the previous example, the team's understanding of the situation that derives from the team situation model can influence both individual situation models and team process. Because individual situation models are tied to individual mental models there is also the possibility for indirect influence of the team situation model on the longer-term shared mental models. Finally, when actions are taken that are based on the team situation model, the outcome of these actions can also serve as feedback to these same factors.

Although this framework for team SA greatly oversimplifies the current body of work on SA and team SA (e.g., each of these factors is itself a complex structure or process), it is intended to highlight those factors associated team SA that seem amenable to measurement. On the other hand, there are some issues tangential to this framework that deserve additional discussion because they are also central to the measurement of team SA.

For example, the term "shared" appears in this framework in the context of shared mental models and shared understanding. Measurement of these factors requires an understanding of what it means for knowledge or a current understanding to be shared. First we assume that a team per se does not have knowledge or understanding, but that there is a body of knowledge (or an understanding) that results from the coming together of individual team members. In this coming together or integration, that there are many variations. For two team members, knowledge can be identical, completely distinct, or at some point in between, though, some have concluded that the identical and distinct extremes are highly unlikely (Rentsch & Hall, 1994). Furthermore, knowledge for two individuals may be similar, or dissimilar and conflicting, or dissimilar yet complementary. For instance, two or more individuals may each have knowledge about a particular topic such as aircraft engine malfunctions. All members of a team may know about the troubleshooting and emergency procedures aspects of engine malfunctions and thus have similar knowledge of the topic. This team would be highly homogeneous in respect to this knowledge. But, some tasks may be best suited for delegation of responsibility and distributed expertise and thus require distributed knowledge or understanding. For example, one individual may know about engine malfunctions and another about emergency procedures and thus as a pair, shared knowledge is complementary. Finally, in another case, two team members who both have knowledge about engine malfunctions may disagree. In this case the knowledge is conflicting. Unfortunately, use of the term "shared" in this context has not discriminated between these variations. These distinctions have implications for the measurement of shared knowledge or understanding in that intrateam knowledge similarity measures reflect only one version of sharing and do not do justice to heterogeneous teams.

One other issue that is relevant to the measurement of team SA has to do with the fact that team SA takes place within an information rich, dynamic, cue-based, and complex multioperator environment. This means that the objects of measurement, such as the situation model, are moving targets and capturing this dynamic property would require at the least, the measurement of multiple situation models over time. Another implication is that team SA cannot be fully appreciated or measured in a task environment that is devoid of most of the contextual elements.

In the following section, current methods that have been used to measure individual SA are briefly described, with the goal of leveraging these measures for team SA or developing new methods directed at the measurement of team SA.

## MEASURES OF INDIVIDUAL SA

Just as conceptualizations of team SA have drawn heavily from theories of individual SA, it is only appropriate that the development of measures of team SA consider the current state-of-the-art in the measurement of individual SA. Various methods for measuring individual SA have been developed (e.g., Endsley, 1990; Fracker & Davis, 1990; Prince, et al., 1992; Sarter & Woods, 1991; Taylor, 1989) and reviewed (e.g., Adams, Tenney, & Pew, 1995; Endsley, 1995b, Fracker, 1991, Salas, et al., 1995; Sarter & Woods, 1995; Taylor & Selcon, 1994; Vidulich, 1989).

The majority of methods fall into four major categories: 1) query methods, 2) performance-based measures, 3) subjective ratings, and 4) behavioral checklists. *Query methods* include those methods that require reports from memory about the details of events in the environment (i.e., the situation model).

Concurrent methods probe memory during task performance, whereas retrospective methods probe memory in a session subsequent to performance. *Performance-based measures* involve having the operator perform a task while measurements (i.e., speed, accuracy, sensitivity based on signal detection analysis) on the overall tasks or subtasks are recorded (Fracker, 1991; Sarter & Woods, 1991). Degree of SA can be inferred from these outcome measures under the assumption that a higher degree of SA leads to better outcomes. Rating scales such as SART (Taylor & Selcon, 1994) and SWORD (Vidulich, 1989) have been used to collect *subjective ratings* (by self or observer) of the degree of SA. In some cases Likert scales are used to report absolute level of SA and in other cases the respondent compares the level of SA from one situation to another (Fracker, 1991). *Behavioral checklists* in which an observer notes any pre-specified SA-relevant behaviors, have also been used to measure SA and team SA (Shrestha, et al., 1995; Brannick, Prince, Prince, & Salas, 1995; Fowlkes, Lane, Salas, Franz, & Oser, 1994).

Criticisms have been raised regarding each of these measurement methods (Endsley, 1990; 1995b; Fracker, 1991) focusing on a variety of issues including the method's reliability, validity, subjectivity, specificity, sensitivity, artificiality, intrusiveness, reliance on memory, and its ability to capture the complexity or richness of SA. Not surprisingly, each measurement method is strong on some of these counts and weak on others and each has produced useful results regarding individual SA.

More important than these criticisms for the purpose of potentially extending these measures to team SA, is the intended scope of these individually-oriented methods in regard to the framework outlined in Figure 1. First, and most obviously, these methods have been developed to measure SA of individuals and not teams. Second, note that even within the purview of individual SA, these methods focus on the situation model or outcome performance and thus, the information that they reveal pertains primarily to the final state or product of SA, rather than the cognition underlying it in terms of the cognitive structures and processes involved in reaching that state. This underlying cognition has been examined to varying degrees outside of the context of SA. Although there is a large literature on the measurement of individual mental models, there is less on the measurement of the situation assessment process or the cues that are used by this process. Finally, and related to the second point, is that the scope of these measurement methods is further reduced by their intended application. They are intended to assess the final state of SA, and in doing so generate single-values indicating degree of SA, rather than a more qualitative look at the content of that state. In sum, the methods for measuring individual SA are primarily suited for the assessment of the final state of SA of an individual.

To be fair, these methods are typically applied to the assessment of individual SA and the evaluation of designs for improving individual SA. For such applications this focus on assessment of the state of SA is on target. However, even at the level of the individual, this outcome orientation is insufficient for uncovering the more detailed information leading to SA that would be required for training interventions or even for informed design decisions (i.e., the content of the situation model, as well as information about situation assessment, cues, and mental models). It is one thing to know for instance, that the operator needs to identify enemy aircraft, but another thing entirely to specify the perceptual cues that are the basis for such discrimination. Training individuals to be aware of enemy aircraft may be more successful if discriminating cues were made explicit (Biederman & Shiffrar, 1987). Note that a focus on the situation model or outcome performance, even if content were identified, would miss these cues and other information that serve as precursors to the situation model. For these same reasons, methods for measuring team SA need to provide procedures for eliciting, as well as assessing, the content of team SA in a context that includes its precursors, as well as its results. Elicitation methods can provide detail that assessment methods lack.

Furthermore, even the *assessment* of a team's SA cannot rely on a simple aggregation of SA scores across team members, but instead requires a deeper look into the content of the team situation model. As suggested in the framework in Figure 1, one predominant feature of team SA is shared knowledge (i.e., shared mental models, team situation model). Without methods that elicit the facts or rules that make up that knowledge, meaningful judgments about shared knowledge are impossible to make. As an analog, just because two individuals obtained similar scores on a classroom test, does not indicate that they share the same knowledge regarding the test items. Similarly, two individuals could have very discrepant situation models, but obtain similar query accuracy scores. Perhaps two individuals (e.g., a pilot and an air traffic controller) both have nearly perfect understandings of the current situation, but they each possess two completely different views or frames of reference. It is clear that a measure of shared knowledge needs to take into account more qualitative information about knowledge content. Although the individual SA measures can be adapted for the purpose of comparing situation models (e.g., item analysis on queries), they have not typically been used in this way. Thus, elicitation methods are especially necessary in the course of measuring team SA.

Finally, methods that provide for elicitation of the content of SA, as broadly defined, can contribute to methods used to assess individual SA. In general, the latter requires assumptions about the content of SA in order to generate the required materials (i.e., queries, performance scenarios, etc.), however, they do not themselves identify this content (i.e., the information in the situation model, the specific facts and rules relevant to mental models, the informational cues in the environment, and the processes involved in interpreting this information). Further, the best intuition or expert judgment regarding construction of these materials may be insufficient in this regard. Research has indicated that expertise is often associated with finely tuned abilities to recognize and classify domain stimuli (Glaser & Chi, 1988), followed by the retrieval of appropriate responses (Klein, 1994). Thus, skill often involves attending to subtle cues or complex patterns in the environment that the expert may herself be unable to verbalize. Thus, the information that is critical for SA is not always obvious, even to an expert. Also the predictive validity of a measure does not ensure that the items that are involved have content validity. For instance, accuracy on queries that target the crew member's ability to identify severe weather patterns may correlate highly with other measures of performance, but this relationship may be attributed to the mutual correlation between a third factor (maybe knowledge of the system) and both weather forecasting and performance. Thus, in this example, training materials that targeted weather forecasting may fail to improve the crew member's performance. This situation may not be problematic until a practitioner decides to go beyond performance prediction and make use of the test items by developing a training program around them. Therefore, elicitation methods can provide useful information for assessment methods. Indeed, elicitation methods are sometimes used in the course of developing assessment methods, but reports of the systematic application of elicitation methods in this area are uncommon.

In summary, the measurement of team SA needs to move beyond the scope provided by methods for measuring individual SA. In addition to targeting teams as opposed to individuals, such methods need to consider SA more broadly by taking into account the aspects of team SA (including individual SA) listed in Table 1 and by providing for elicitation as well as assessment. The elicitation of team SA as broadly conceived in Figure 1 will provide information necessary for training and detailed design decisions, will provide for a more precise measurement of shared knowledge, and will augment current assessment methods. In the following section we describe a knowledge elicitation approach to measurement that targets these requirements.

## APPLYING KNOWLEDGE ELICITATION METHODS TO THE MEASUREMENT OF TEAM SA

Knowledge engineering is the process of building a knowledge-base, required for intelligent software such as expert systems, intelligent tutors, and decision aids. One approach to knowledge-base creation has involved eliciting knowledge from a human domain expert. Within this context, the goal has been to reveal the domain facts and rules (i.e., content) relevant to a particular task, a goal that has become more central for cognitive engineering in general and that echoes some of our objectives for the measurement of team SA. More detailed definitions of knowledge elicitation, along with historical development of the enterprise and associated issues can be found in Cooke (1994), Cooke (1999), and Hoffman, Shadbolt, Burton, and Klein, (1995).

## OVERVIEW OF KNOWLEDGE ELICITATION METHODS

A wide variety of knowledge elicitation methods has been applied to the elicitation of knowledge at the individual level. Most of these methods fall into one of four categories : 1) observations, 2) interviews and surveys, 3) process tracing, and 4) conceptual methods. Each category contains a number of methods that vary in terms of the specific procedure and the type of knowledge that is targeted. Whereas details can be found in the aforementioned reviews, a brief overview of each category of knowledge elicitation methods follows. We propose that the measurement of team SA can benefit from leveraging these knowledge elicitation methods.

Observations can take written, audio, or video forms and have the potential to provide vast amounts of information in the context of performance, while interfering minimally with tasks. Variations within this category revolve around issues of how to observe (passively or actively) and what to observe (predetermined features or specific aspects of the task and environment). The advantage of minimal intrusion is offset by difficulties interpreting the data. Recently, however, tools to aid in this process such as video analysis software have been developed. There are tools available that interface video recorders with personal computers so that annotations, counts, and classifications of behavior can be entered as the video is viewed (Harrison, 1991). Then, if desired, video clips corresponding to a specific type of event can be automatically located and played back. Also, various analyses of frequencies and transitions are provided. Timelines (Owen, 1993) and MacSHAPA (Sanderson, McNeese, & Zaff, 1994) are two tools with these capabilities.

Interviews and surveys, like observations, are excellent for gaining a general understanding of the situation, and generating and verifying hypotheses. Unstructured interviews are free-form in that neither content nor sequencing is prespecified. Structured interviews follow a predetermined format and can vary from highly rigid to only loosely constrained. Although structured interviews require more preparation time than unstructured interviews, they also have advantages of being more systematic, and therefore more complete and comfortable for both participants. With enough structure, a structured interview can be converted into a paper-and-pencil measure, such as a survey or questionnaire, which is often easier to administer than interviews.

Process tracing techniques are methods for collecting data concurrently with task performance. These data are later analyzed to make inferences about the knowledge or cognitive processes underlying task performance. One of the most popular forms of process tracing relies on verbal reports provided while the participant is "thinking aloud" during task performance (vanSomeren, Barnard, & Sandberg, 1994). These reports can be retrospective, but it is typically believed that as long as there is no interference with the task itself, concurrent reports provide more valid information regarding the current contents of working memory. Nonverbal data including keystrokes, actions, facial expressions, gestures, and general behavioral events have also been collected to trace cognitive processes.

Protocol analysis is the term used to describe various methods for summarizing and interpreting process tracing data (Ericsson & Simon, 1996). It typically involves transcribing the data, developing a coding scheme that captures the critical content of the data, applying the coding scheme to each identified unit in the protocol, and exploring frequencies, patterns, and sequential dependencies in the results. A relatively new area of inquiry that focuses on ESDA (Exploratory Sequential Data Analysis; Sanderson & Fisher, 1994) has produced some analytical tools and methods in recent years to facilitate this process. The combination of these new tools and event data captured on-line has promise as a way of overcoming many of the costs associated with knowledge elicitation.

Conceptual methods produce representations of domain concepts and their relations. Methods included in this set include cluster analysis (Johnson, 1967), multidimensional scaling (Shepard, 1962a; 1962b), Pathfinder (Schvaneveldt, 1990; Schvaneveldt, Durso, & Dearholt, 1989), and concept mapping (Jonassen, Beissner, & Yacci, 1993). In general, the methods take pairwise estimates of relatedness for a set of concepts and generate a spatial or graphical representation of those concepts and their relations. The general goal of these methods is to reduce the set of distance estimates in a meaningful way. Resulting representations can then be compared qualitatively and in many cases, quantitatively across groups and individuals. The techniques tend to be indirect in that they require judgments about conceptual relatedness, as opposed to introspections or explicit verbal reports. An additional advantage of these methods, particularly for measuring team SA, is that they can handle results from multiple individuals, including summarizing data from several individuals and comparing data across individuals or groups.

Although knowledge elicitation methods have been used extensively in research on individual mental models (e.g., Rowe, Cooke, Hall, & Halgren, 1996) and to a lesser extent as a front-end to the assessment of individual SA (e.g., Endsley, 1989; Endsley, 1993; Endsley & Rogers, 1994), there has been limited research to date in which these methods have been applied to the measurement of team SA. In a few cases, process tracing methods have been applied to team SA, focusing specifically on communication events. For example, the frequency of cue-specific communication has been used as an index of team SA. Orasanu (1990) has found a positive relationship between SA, as measured in terms of communication events and performance. Similar work examining communication and team performance has been conducted by Mosier and Chidester (1991) and Bowers, Braun, and Kline (1994). There have similarly been several applications of conceptual methods to the shared mental model aspect of team SA (Blickensderfer, Cannon-Bowers, & Salas, 1997b; Gualtieri, Fowlkes, & Ricci, 1996; Klimoski & Mohammed, 1994; Kraiger, et al. 1996; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000; Stout, 1995; Minionis, Zaccaro, & Perez, 1995; Walsh, Henderson, & Deighton, 1988). In general these investigators have extended this method to the team level by postulating that assessments of the degree to which knowledge structures of team members are similar can be used as an index of the degree of shared mental models present among team members (e.g., Blickensderfer, et al., 1997b; Kraiger & Wenzel, 1997, Stout et al., 1996).

There are some limitations associated with these methods that may partially explain their limited use. For instance, they can be time-consuming to carry out. On the other hand, in light of targeting the content of team SA, broadly defined, there are no quick methods, but there are systematic ones. Also, methodological developments in this area may ultimately result in new methods that will automate or streamline this approach. Another drawback of these methods is that because expert knowledge is domain-specific (Glaser & Chi, 1988), the results from one domain may not generalize to another. However, though specific results may be domain-specific, it is generally believed that methods are not. Finally, for the most part, these methods have been used to measure knowledge at the individual level. In order to apply them to the measurement of team SA, it is necessary to modify them for team applications. As described previously, steps have already been taken to begin to apply process tracing and conceptual methods to communication and shared mental model aspects of team SA, respectively. In what follows we prescribe a methodology for applying knowledge elicitation methods to team SA as broadly conceived in Figure 1.

## A PRESCRIBED METHODOLOGY FOR MEASURING TEAM SA

Acknowledging that team SA begins with the SA of each individual team member, the methodology that we prescribe integrates both the individual SA and team SA factors listed in Table 1, resulting in four major measurement targets which include: 1) situation models -- those of individual team members, as well as the team situation model, 2) knowledge or mental models -- those for each team member, as well as the knowledge that is shared by team members, 3) cues in the environment and the situation assessment process that individual team members use to interpret those cues, and 4) team process behaviors (see also Stout, et al., 1996). In what follows we prescribe a methodology that applies knowledge elicitation methods to each of these targets.

Each of these targets and the prescribed methods for eliciting this information are listed in Table 2 and discussed in the sections that follow. The purpose of Table 2 and the associated discussion is to propose an agenda for tackling the measurement of team SA. The proposed methods are not new, but have been applied and evaluated mostly at the individual level and in the context of problems other than SA. In general, they need to be applied, evaluated, and in some cases modified, for the measurement of team SA.

[Insert Tables 2a and 2b about here]

The two sections of Table 2, elicitation and assessment reflect the temporal sequence of the general measurement approach that we propose. As emphasized previously, we view knowledge elicitation as central to the measurement of team SA. Elicitation methods generally result in the cues, facts, rules, and strategies that make up the content of team SA and whereas this information alone is useful for applications in training and informed design and as a front-end to assessment, it is essential for investigations of team SA.

On the other hand, assessment methods can and should also be used to confirm that the information elicited and used to derive them is valid. This is important because despite the best efforts of the experimenter to elicit complete and unbiased information, there are no guarantees. The information provided may be ambiguous and the experimenter can make errors of interpretation. Thus, before applications or assessments are based on this information, it is important to confirm the validity of the elicited information. In general, confirmation involves following elicitation with assessment and comparing the resulting assessment metric to other measures of the construct or to changes in individual or team performance. If the information elicited (and then assessed) accurately reflects team SA, then variations in the elicited information should correspond to variations in team performance and to other measures of team SA.

Thus, in what follows we describe the application of elicitation methods to the measurement of team SA, but additionally show how these methods can be tied to the assessment process. The underlying proposition is that the selection of an elicitation or assessment approach should depend on the purpose of measurement, with team SA applications being inextricably dependent on elicitation.

Furthermore, the methods that are prescribed in Table 2 tend to be general categories of methods (e.g., structured interview, process tracing), rather than specific techniques or tools (e.g., critical incident technique, multidimensional scaling). These categories are prescribed on the basis of the type of content typically elicited. For instance, process tracing methods tend to elicit process information, whereas conceptual methods tend to elicit product information and are thus, best suited for eliciting team SA process (e.g., team process behaviors) or team SA products (e.g., shared mental models), respectively. Selection criteria for specific techniques within these categories involve consideration of not only the output of the technique, but also the pragmatics of the procedure involved. See Hoffman, et. al., (1995) for more information on technique selection. However, especially in regard to measuring team SA, more important than the specific elicitation technique used, is the context in which it is applied.

With the exception of mental models and shared mental models which are thought to be knowledge structures that exist in long-term memory, it is important that measurement occurs within the task context. This is because team SA not only occurs in a rich and dynamic environment, but it depends on many of the cues and information in the environment. Interestingly, it has been found that this context dependency is even more critical for situation assessment by experts than novices (Federico, 1995). Similarly, the elicitation of information relevant to team SA will depend on the surrounding context. Thus, before elicitation is even attempted, it is necessary to identify the context and task within which measurement will occur.

Although measurement could occur in a field setting, it is more likely that most studies would occur in a semi-controlled setting such as a synthetic task environment. Context can therefore be provided via high or low fidelity simulations or even paper-based scenarios or retrospective videos. For the purpose of elicitation, the objective is to draw the participating individual or team back into the original scenario so that knowledge is more readily available. It is necessary to know a fair amount about the task and setting to construct the simulated tasks and scenarios. Observations of teams in field settings are ideal for generating valuable information (general procedures, decisions involved, constraints, information requirements) for the creation of synthetic tasks scenarios that are faithful to the actual task, as are more formal task and cognitive task analysis techniques (Kirwan & Ainsworth, 1992). In the discussion that follows it is assumed that the context of measurement has been identified as a prerequisite for knowledge elicitation.

In sum, the prescribed methodology in Table 2 targets four aspects of team SA that are central to its measurement. Some of the targets involve individual as well as team products or processes. Categories of elicitation techniques are suggested for each of the targets along with procedures for using the resulting information for assessment or confirmation purposes.

*Measuring Individual and Team Situation Models*

Most of the measurement work in the domain of individual SA has focused on the individual situation model and the assessment of the degree of SA by identifying accuracies or inaccuracies in information thought to be relevant to the situation model. Queries methods such as SAGAT (Situation Awareness Global Assessment Technique), a concurrent method in which the display is frozen repeatedly during a simulation and queries are randomly presented (Endsley, 1990) and SPAM (Situation Present Assessment Method), a truly concurrent method in the sense that queries are presented while the task is ongoing and all of the information is present (Durso, Hackworth, Truitt, Crutchfield, Nikolic, and Manning, 1998) have often been used to this end.

In this section we prescribe a methodology for measuring team situation models that is focused more on the elicitation of the content of SA than previous individually oriented approaches. In accord with the framework in Figure 1, individual situation models of team members are precursors to team mental models and so are targeted here as well. We begin by describing the methodology as directed toward individual situation models and then extend these ideas to the team situation model.

Information pertinent to the situation model can be elicited through structured interviews with individuals. In fact, in many cases, individual SA measures use some form of interview with one or more experts as a means of developing assessment materials (e.g., Endsley, 1989; Endsley, 1993; Endsley & Rogers, 1994). The centrality of elicitation in our measurement scheme suggests that a greater emphasis and degree of systematicity be placed on these front-end procedures.

There are a number of different ways to structure an interview (Cooke, 1994). For instance, it can focus on past critical incidents, or system diagrams representing the function and interrelationships of components, or a description of what is happening in a current scenario or mission. It is exactly this imposition of structure, which distinguishes responses to structured interview questions from think-aloud verbal reports (to be addressed later), which are thought to provide an indication of ongoing cognitive processes underlying performance. Responses to structured interview questions do not reflect ongoing cognitive processing, but can address specific questions or target areas of concern--situation models in this case.

During the structured interview, the interviewer should prompt the participant for information about her current understanding of the situation and like observations, responses to prompts should be recorded in written, audio, or video form. Questions such as "tell me what is going on in the situation now," "tell me what is about to happen in this situation," and "tell me how the current situation differs from the previous one" would all get at this kind of information. In constructing structured interview questions, it is typically a good idea to begin with rather broad questions and to gradually increase the structure by questioning about specific aspects of the situation. The point is to attempt to elicit accurate and complete information, while at the same time avoiding leading or biasing questions. Paper should be provided during the structured interview so that the interviewee can make lists or even draw diagrams representing elements of the situation and their interrelations. In particular, a concept mapping procedure such as COGENT (Cognitive Engineering Network Technologies; Sanderson, et. al., 1994) in which the participant expresses his knowledge in terms of a concept map in which concepts are represented as nodes and relations between concepts are represented as links between nodes may prove fruitful (e.g., Federico, 1997)

Because situation models only make sense in the context of a situation, it is clear that these interviews need to occur in context (either concurrent with performance, or retrospective with context provided by a video of performance). On the one hand, interviews that occur concurrent with performance may interfere with that performance, but retrospective interviews may be subject to problems of memory retrieval on the part of the interviewee. In this case a video-aided retrospective interview may provide the optimal solution. The choice should be guided by the extent to which each of these problems seems to be relevant.

If concurrent interviews are used then, along the lines of typical SAGAT procedures, the task could be frozen just prior to the administration of questions.

Similarly, following the SPAM methodology, the task could continue while the operator answers questions, focusing particular on future events. However, it is important to distinguish the structured interview questions from SAGAT queries. SAGAT queries are typically more specific and less open-ended (e.g., "tell me the location of the enemy aircraft") than structured interview questions. Also, queries typically assume the contents of the situation model and are used to assess its completeness and accuracy. Structured interview questions are more open and intended to reveal the contents of the situation model, rather than assess it (e.g., "tell me what is important in the situation at this moment, " or "tell me the important features of that aircraft in this situation").

The outcome of the structured interviews is a compendium of responses to specific questions or prompts. These responses will only be as good as the prompts that elicit them, so it pays to carefully construct interview questions. Whereas, raw responses to prompts may be very informative with regard to constructing queries or training program design, if a number of data sets are collected across different individuals or scenarios, there are a variety of ways that the interview responses can be summarized. Although, interview responses are not the same as think-aloud verbal reports, this does not preclude categorization and summarization of the responses along the same lines as coding of verbal reports associated with protocol analysis. For instance, in response to the question, "tell me what has changed in the situation in the last five minutes" there are most likely a finite number of different responses that occur across different interviewees in any one scenario. Categories of responses can be initially derived by judges using a subset of the data and refined as new data are analyzed. Once the various categories are derived, individual responses can be coded accordingly and summarized across interviewees. Note that the identification of categories is an interpretive process that is usually guided by some theory or framework concerning the types of elements in a situation model for that task, although the theory may be implicit.

Table 2 points to query methods as appropriate for assessing situation models at the individual level. This suggestion is not new, as query methods have been commonly used to measure individual SA of the type targeted by the situation model. What is new is the elicitation emphasis, which suggests that queries be constructed on the basis of the results of systematic elicitation. So, for instance, if the interviewee responds that she believes that in a particular situation that there is a weather disturbance that necessitates abandoning the mission, then this aspect of the situation model can be transformed into a query in which another participant is asked more directly about the weather situation. Note again, that the resulting queries should be much more specific than the original structured interview prompts. A common outcome of this method is a score that reflects query accuracy. This can be used to confirm elicited data by correlating query accuracy scores with other measures of performance. If the queries reflect SA, and if SA is critical for performance, then query accuracy should correspond to performance. Assessment of degree of SA can then be based more confidently on query accuracy.

This discussion of situation models has thus far focused on eliciting and assessing the individual situation model. How can these methods be extended to the measurement of team situation models? The most obvious approach is to compare the individual situation models across team members. Of course, such comparison has to be tempered with the understanding that the team situation model, while partly dependent on the individual situation models, is more than this simple summation of parts. With this caveat in mind, the critical question is one of comparing situation models within a team.

If we assume that individual structured interviews elicit information concerning the situation model, then it is the structured interview output that requires comparison. The first pass at this comparison should identify the range of responses to each question and establish some preliminary criteria for determining whether answers are the same, or different and if different whether they are similar, dissimilar and conflicting, or dissimilar and complementary. At this point, any categorization done in the elicitation phase should facilitate such a comparison. Also, the less structured the interview, the more difficult this comparison will be, as participants may not address the same issues. One goal of such a comparison is to aggregate the responses to identify the team's understanding of the situation. Thus at this level, the product of elicitation is a summary of this comparison process (e.g., 60% of items mentioned in the situation model were the same or similar across team members; there was one conflict in the situation models for 2 of the team members). This summary should capture the team situation model by identifying specific features of the situation model for which there are similarities or discrepancies and the nature of the discrepancies (i.e., complementary or conflicting information). See Cooke, Salas, Cannon-Bowers, and Stout (2000) for suggested metrics for comparing and summarizing intrateam knowledge differences.

At the team level the assessment step would involve generating queries that specifically target interesting team situation model features such as conflicts among team members. These team-oriented queries could be added to the list of queries derived from individual elicitation. The generation of a metric representing the team situation model would require a query comparison procedure similar to that performed on the structured interview responses, however, responses to the very specific queries should be easier to compare than responses to the more general structured interview questions. It is important, however, to point out that query comparison is not simply a matter of comparing accuracy scores among team members, because two individuals could be very accurate with complementary views. Instead, it involves doing a content analysis of the query responses and comparing that analysis across team members.

The outcome of this comparative analysis on the queries should provide an indication of whether the specific information in the team situation model that is associated with the queries is similar or dissimilar across team members, and if dissimilar, whether conflicting or complementary. For instance, it may be determined that for a particular aviation task with a high performing three-person crew that two of the three team members have common situation models, whereas, the third member holds a view of the situation that is complementary to the other team members. The result of this comparative process can also be quantified in terms of proportion agreement or team accuracy. Then, accuracy or similarity metrics at the team level can be correlated with team performance to determine if and precisely how shared situation models are related to team performance. So, for example, it may be the case that teams that have situation models that are more similar to one another also have high performance scores compared to teams that have dissimilar situation models. Once confirmed, then this same information can be used in a predictive way (e.g., teams that have similar situation models would be predicted to also perform better than teams that have dissimilar models).

In short, the process that we propose for measuring the team situation model involves a detailed content analysis of responses to structured interviews and queries among individuals within the same team. The former information should elicit content relevant to the team situation model, and the latter should provide an assessment of the accuracy of the team's situation model and the similarity of the model among individuals on the same team.

*Measuring Mental Models and Shared Mental Models*

The second targeted component of team SA is knowledge in the form of individual and shared mental models about the task, system, or team members. Mental models are part of long-term memory, as opposed to situation models, which are constructed as needed. Knowledge elicitation methods that were originally intended to measure long-term knowledge of individuals have been extensively applied to the measurement of individual mental models (e.g., Coury, Weiland, & Cuqlock-Knopp, 1992; Gillan, Breedin, & Cooke, 1992; Gray, 1990; Kempton, 1986; Payne, 1991; Rasmussen & Jensen, 1974; Vosniadou & Brewer, 1992) however, not typically in the context of individual SA. Only recently have investigators attempted to measure this knowledge at the team level (i.e., shared mental models; e.g., Gualtieri, et al., 1996; Kraiger, et al., 1996; Mathieu, et al., 2000; Stout, 1995; Minionis, et al., 1995; Walsh, et al., 1988). In this section we identify some methods for eliciting and assessing mental models and shared mental models which draws heavily from the existing work in this area.

Because it is assumed that team members come to the task with mental models, as opposed to situation models, which they form during task participation, mental models can be elicited apart from the actual task context. However, it is also assumed that the team members have had past experience with identical or similar systems, tasks, and team members in order to have developed these knowledge bases over the long term.

In Table 2 both structured interviews and conceptual methods are prescribed for eliciting mental models. (Note that this is just one example where a combination of two or more KE methods or method categories can be successfully applied to the same problem). Whereas interviews to elicit situation models are focused on the specifics of the situation, structured interviews can be focused on long-term knowledge in several different ways, depending on the type of knowledge that is to be elicited. Knowledge about a system can be elicited by having each team member draw a diagram representing system components and then having her draw directed links between components representing functional relations. Mental models about the task can be elicited by starting the interview with an outcome or goal (e.g., abort landing) and asking the team member to work backwards from that outcome by listing the information that would have led to such a decision, followed by the information or cues that would have preceded that information and so on. Finally, mental models about team members can be elicited through two structured interview techniques. One involves having the team member draw a diagram with nodes representing individual team members and links between nodes representing information flow between team members. Another technique that could be used here is a structured interview that centers on role reversal. That is, a team member assumes the position and duties of another team member during a simulation. The interview can then focus on views of the task and system (and situation, if engaged in one) from the other team member's position.

These structured interview techniques represent a subset of the techniques that can be used to elicit the relevant long-term knowledge of individuals (see Cooke, 1994 for additional suggestions). The decision to use one or more specific techniques will undoubtedly hinge on the nature of the team task being analyzed. For instance, in some cases, it is not critical that the operator have a complete understanding of the system, so this type of knowledge may be overlooked. Likewise, some tailoring of techniques to meet the specifics of the task may also be required. As suggested for situation models, the responses to the structured interview questions can be used in their raw form or categories can be generated and responses coded as discussed previously. Thus, the product of this method is a set of raw or categorized responses to interview questions.

Conceptual methods are also ideal for eliciting information associated with mental models. These methods include multivariate techniques of cluster analysis, multidimensional scaling, and Pathfinder network scaling and each produce representations of domain concepts and their relations that depict the underlying structure of some subset of knowledge. The structured interviews serve as a more open-ended method for eliciting mental models and can serve as a front-end to the more directed conceptual methods. For instance, the categories generated to code the structured interview responses should provide a good source of concepts or items central to the domain of interest that are required by conceptual methods. After a set of concepts is identified, the next step involves the collection of relatedness estimates for all pairs of concepts from each team member. Depending on the number of concepts involved, individuals could sort the items into meaningful categories (more than 30 concepts) or rate them for relatedness (30 or fewer concepts).

As an aside, whereas structured interviews are suited for elicitation of either situation model- or mental model-type knowledge, conceptual methods are ideal for eliciting mental models, but would be highly intrusive if used to elicit situation models because the task would need to be interrupted long enough to collect a hundred or so relatedness ratings. Recent research has taken steps toward solving this problem by attempting to elicit the relatedness information on-line during task performance. Specifically, patterns of behavior during performance are matched *a priori* with knowledge structures so that ultimately the knowledge structure can be inferred from the pattern of behavior (Rowe, et al., 1996). At any rate, task intrusiveness is not a problem for eliciting mental models because the knowledge in long-term memory should not be altered drastically by performing the rating task.

Once relatedness estimates have been collected, they can be submitted to one of several algorithms such as Pathfinder network scaling (Schvaneveldt, 1990). This procedure not only reduces the relatedness estimates to a more manageable and interpretable result, but these results can then be compared qualitatively and quantitatively. Although relatedness ratings, themselves, can also be compared quantitatively using correlation coefficients, research has indicated that Pathfinder network results reduce the data in a way that is more psychologically meaningful than ratings or multidimensional scaling (Cooke, Durso, & Schvaneveldt, 1986; Cooke, 1992a). Thus, the product of conceptual methods is a graphical representation (e.g., network, spatial layout) of the individual's knowledge structure for the selected set of concepts, along with indices (e.g., spatial coordinates, dimension weights, pairwise distances between concepts) that can be derived from such a structure.

As previously mentioned, one advantage of conceptual methods, particularly for measuring team SA, is that they can handle results from multiple individuals, including summarizing data from several individuals and comparing data across individuals or groups. These procedures are precisely what are needed for assessment. For example, there are methods to quantitatively compare resulting representations that are very much like correlational measures. In the case of Pathfinder, any two resulting networks can be compared in terms of the proportion of shared links (i.e., the network similarity measure, Goldsmith & Davenport, 1990). In fact, methods like these have been used most often to identify group or individual differences in cognitive structures (e.g., Goldsmith, Johnson, & Acton, 1991; Rowe, Cooke, & Rivera, 1998). Thus, these methods can be used, for instance, to compare the knowledge structures of a pair of individuals to determine degree of compatibility. The output of conceptual methods can also be compared qualitatively across such dimensions the basis of presence or absence of links for specific pairs, most highly central nodes, and overall complexity of the node-link structure. The availability of quantitative and objective intrateam comparison methods for the results of conceptual methods is a striking advantage over the cumbersome comparison process required for structured interview results. Thus, in Table 2, the assessment methods for mental models include the former, but not the latter.

The ability to compare these conceptual representations enables a host of other analyses. First, by way of confirmation of the information obtained in the elicitation procedures (i.e., structured interviews, conceptual methods) it can be determined whether variations in the networks resulting from these methods are associated with performance variations. This can be done through the creation of a referent network structure against which each individual's network is compared. The referent could be the average or aggregate (assuming sufficient intragroup agreement) of the networks of the top performers in the sample (or external to the sample) or it could be derived logically with the help of experts. It is expected that if the techniques described above are capturing mental models, then the proportion of links shared with the referent structure should correlate with performance measures. Not only can this metric (accuracy based on similarity to referent structure) be used to confirm the elicitation results but once confirmed, it can be used to assess the mental models of individual team members. Those who have structures more closely resembling the referent should have higher performance scores than those whose structures are dissimilar from the referent.

In terms of adapting this methodology to team SA, the ability to compare two conceptual representations is again fortuitous. For any team, the mental models of all pairs of individuals can be compared qualitatively (at least for smaller teams) and quantitatively as described above. Note that the quantitative comparison will only reveal degree of similarity (i.e., proportion of shared links, correlation of distances) and not the content of the information shared. These details require a more thorough qualitative analysis of the representations such as an identification of link labels in Pathfinder networks (Cooke, 1992b). There are also ways to derive a team knowledge representation that represents the knowledge of the team as a whole. Give adequate similarity among team members an aggregate representation could be derived by averaging the proximity data across team members. But in some cases, in which there is minimal intrateam similarity among conceptual structures, this aggregation procedure could hide interesting individual differences. Alternative approaches to aggregation include averaging data across the subset of team members who have high intragroup similarity or embellishing representations with agreement information (e.g. weighting links in a network in accord with the number of team members who have this link in their individual knowledge structures). Thus, the product of the team elicitation is a representation of team knowledge, guided by an understanding of the extent of intrateam similarity.

To confirm this information, the variations in intrateam similarity among teams used in the elicitation stage to aggregate individual knowledge structures, can be compared to team performance measures. If shared knowledge is critical for team performance and if the similarity among individual knowledge structures reflects the degree of shared knowledge, then intrateam similarities should correspond to team performance measures. In addition, the accuracy of the shared mental model can be assessed by comparison of the aggregate knowledge structure or structures to a team knowledge referent. Again this referent could be either logically constructed or empirically derived from high-performing teams. As is also the case for assessing individual accuracy via comparison to a referent, the accuracy metric will only be as good as the referent is a meaningful standard. See Rowe, et al. (1998) for issues concerning referent selection. An accuracy metric can be based on the similarity between the team referent and the team knowledge structure and should indicate the degree to which the team knowledge structure is accurate which if psychologically valid, should correspond to team performance. Once the methodology and its results are confirmed by finding a relation between both intrateam knowledge structure similarity and shared mental model accuracy and team performance, these methods and metrics can also serve an assessment function.

Note that one general advantage of the elicitation approach advocated here is that the metrics are grounded in rich qualitative data, which have the potential to serve a diagnostic function. For instance, a low team accuracy score could be followed by a more detailed analysis of similarities and differences between the team and referent knowledge. In this analysis potential misconceptions or areas of missing knowledge for each team or individual can be identified, which can serve as a target for intervention. See Cooke and Schvaneveldt (1988) and Schvaneveldt, Durso, Goldsmith, Breen, Cooke, Tucker and DeMaio (1985) for examples of this kind of diagnostic function of knowledge structures as applied to individuals.

*Measuring Cues and Situation Assessment*

According to the framework described earlier, situation assessment of environmental cues occurs at the individual level, but is instrumental in the formation of the situation model that is relevant at both individual and team levels. Thus, the knowledge elicitation methods described in this section, target cues and situation assessment for each individual team member.

Cues and the processes operating on them are probably the most difficult information relevant to SA to elicit, because although individuals may be aware of categorizing a situation, skilled operators often process the cues that led to a particular category automatically, without conscious awareness. Furthermore, constellations of cues that allow for more holistic processing of the stimuli in the environment (for example entire chess positions, Chase and Simon, 1973a; 1973b) make the job of explicating the cues even more challenging. However, in the case of cues, it is clear that their elicitation, as well as the elicitation of the processes used by individuals to interpret them, has to occur in the context (actual or synthetic) of the actual task.

Video taped observations of a team at work may generate some preliminary ideas about the informational cues in the environment affecting the situation model of each team member. At the very least, these kinds of data will serve to rule out some cues to which the individual is not attending and to identify information to which the individual directly attends as potential cues. Due to the complex nature of tasks involving team SA, it may be necessary to have multiple observers and/or video cameras. The observers and cameras should be located such that they are as unobtrusive as possible. An observation booth behind a one-way mirror would be ideal. It is also typically the case that soon after experimental participants are informed of the video cameras, they forget about them and resume the task as usual. Finally, because team SA tasks typically require a great deal of domain expertise, it may be necessary to have experts retrospectively interpret the observations with the aid of videos. In any case, the video analysis software discussed previously would also be helpful here. The product of the observations is hypothetical cues and noncues.

A slightly more directed approach to cue elicitation is to ask individuals or teams to categorize stimuli. The co-occurrence of stimuli within the same category provides an indication of relatedness that can serve as input to conceptual methods. Methods such as multidimensional scaling (Shepard, 1962a; 1962b) are particularly useful in identifying the dimensions that underlie data of this kind. The dimensions themselves can then be interpreted as hypothetical cues, the product of this method. For instance, Federico (1995; 1997) used a sorting and labeling procedure, in which scenarios were sorted into groups and various statistical analyses were used to identify the cues that were used in this sorting. Note that although more directed, this method relies on the comprehensiveness and representativeness of the original set of stimuli to be sorted. Omission of a few critical stimuli may result in omission of one or more critical cues. On the other hand, the data are much more easily interpreted compared to the less directed observational techniques.

To elicit the situation assessment process, methods of process tracing can be applied by having the individual think aloud during the assessment process itself (or retrospectively in conjunction with viewing a video tape). These verbal reports should reflect the cognitive processing underlying task performance, at least in terms of the current contents of consciousness. The procedure of verbal reporting does not always come easily and may be facilitated for the participants by using dyads or larger groups in which two or more individuals think aloud to each other. This practice lends itself well to the elicitation of team situation awareness in which there are multiple knowledgeable parties.

Once the data have been collected, the resulting protocol should be analyzed in order to identify critical events in the situation assessment process. This involves generating a code that captures the variations in the protocol. Note that the generation of such a code is typically theory-driven and thus, top down in nature. This means that the situation assessment information (cues, cue categories, cognitive processes) that is the output of this process may be biased in favor of the theoretical framework underlying the protocol analysis. However, like the products of other elicitation methods, the process tracing output is no exception to the confirmation requirement.

Once the code is generated, the protocol needs to be segmented and segments need to be categorized in accord with the code. To facilitate this process, several protocol editing and analysis tools have been developed in the ESDA tradition including for instance, MacSHAPA (Sanderson, et al., 1994). In addition, PRONET (Cooke, Neville, & Rowe, 1996) is a procedure for generating node-link graph structures from the co-occurrence of items in a sequence. Unlike the conceptual methods that produce graph structures of concepts and their relations, PRONET results in graph structures representing procedures in terms of event or behavioral sequences. In the case of process tracing, those items would be units of verbal reports or nonverbal behavior.

There are several methods that can be used to verify the accuracy of the cue information elicited. If the cues are visual in nature, they can be confirmed (or not) with eye movement data. In general, nonverbal data can also be used to trace cognitive processes. These include eye movements, as well as keystrokes, actions, facial expressions, gestures, and general behavioral events. Eye movements in particular have been used to identify cues used by experts (Carmody, Kundel, & Toto, 1984) and even experts in the context of pilots' mental models (Wickens, Bellenkes, & Kramer, 1995). Despite the fact that eye movement instrumentation is often restrictive and therefore somewhat intrusive, the data it generates are useful for getting at order and weight information in sequential tasks, as well as for clarifying and verifying (Lauwereyns & d'Ydewqalle, 1996). Although there is not a one-to-one mapping between eye movements and thoughts, they can be used in conjunction with verbal reports to help to verify them. For example, if an operator claims to be attending to a particular cue, but never scans the screen where that cue is located, then the accuracy of the verbal report would be called into question. The lack of visual attention paid to a particular cue is stronger evidence that the cue is not used, than is positive evidence of visual attention to a cue.

In other cases, confirmation can be accomplished by systematically varying the values of the hypothesized cues in a set of scenarios and regressing action or decision outcomes on these values (i.e., policy capturing). High or statistically significant regression weights associated with cues support the hypothesis that the cue is used. Another way to verify the information elicited regarding situation assessment is to alter the scenario or task in such a way as to prevent situation assessment. If the elicited information is accurate and if situation assessment is as critical to SA as assumed, then overall task performance should suffer when situation assessment is prevented. This technique is subject, however, to the same criticisms raised earlier in regard to artificial task manipulations.

The same procedures used to confirm the cue information elicited can later be used to assess individual SA in regard to cues and situation assessment. Eye movements can be monitored to determine if critical cues are ignored or if some cues are over- or underemphasized. Policy capturing can identify the extent to which each cue plays a role in an individual's action or decision compared to a set of referent weights obtained empirically or logically. Finally, scenarios can be constructed so that individuals who use cues effectively will perform in a predicted manner.

In sum, the elicitation of cues and the situation assessment process is central to individual SA and thus plays a role in individual situation models. The elicitation methods described in this section can help to suggest a set of potential cues and then using assessment methods the validity of this information can be evaluated. Although challenging, the process of understanding the environment from the perspective of an experienced operator can provide a number of opportunities for interventions in training and design.

*Measuring Team Process*

Team process behaviors by definition, involve multiple team members, and therefore need to be measured at the team level in the context of the task. In the context of team SA, team process is thought to be responsible for the integration of individual situation models and shared knowledge to result in the team situation model. Thus, team process is to team SA what situation assessment is to individual SA.

Several investigators have measured team process, focusing on communication events--the types of events, and their frequency (Bowers et al., 1994; Orasanu, 1990; Mosier & Chidester, 1991). In general, this research has revealed a relationship between communication processes and team performance. Although good team process is probably linked to better team SA, as well as performance, the exact nature of the team process behaviors that result in effective team SA in any given task and situation require methods that can elicit this information.

Elicitation techniques can be used to broadly elicit team process behaviors in general within a specific context. In particular, the process tracing methods discussed above in the context of situation assessment are ideal for interpreting patterns of complex events and behaviors in terms of more abstract behavioral patterns or the cognitions that underlie such patterns. Instead of verbal reports as was suggested for situation assessment, the process data collected might consist of written observations, computer event logs, or video records of team behavior. In the analysis of this protocol, specific team behaviors can be identified (again a theory-driven process) and then used to code the remaining protocol. For instance, the types of communication events observed should be made explicit (e.g., nonverbal communication, command, request for information, etc.). Note that although protocol analysis has a significant top-down component, it is nonetheless the case that the protocols themselves provide a rich source of bottom-up information. Some component of empirical derivation of the categories is important, given that the goal is to identify the content of team SA.

Beyond the frequencies with which categories or transitions between categories occur, techniques such as PRONET (Cooke, et al., 1996), can be used to help identify global behavioral patterns. This technique involves calculation of co-occurrence frequencies or transition probabilities for all pairs of events and behaviors noted. These values are then submitted to a multivariate procedure such as Pathfinder. Pathfinder will reduce the set of probabilities or frequencies to a node-link structure in which events are represented as nodes and frequently occurring transitions between events as links between nodes. See Cooke, et al. (1996) for an example of this procedure. Using this kind of graphical representation of team behavior can, for instance, lead to hypotheses about sequential dependencies among the behavioral events.

This same procedure can be used to analyze interactions among team members, where events are coded not only by type, but also by team member (i.e., who is talking to who). In this case between-element transitions would represent the frequency of communication for each pair of individuals. Note that this is different from the proposed method of eliciting mental models about team members in which an individual's knowledge about team information flow is elicited. In the case described here, the actual behavior itself is recorded and analyzed. In the same way, behavior relevant to team coordination, planning, or leadership can also be recorded and analyzed. The product of this analysis is an abstraction of the behavioral events in terms of a protocol, frequencies, and a procedural network.

To evaluate the validity of the process tracing results, differences between the frequencies or resulting procedural representations can be correlated with team performance differences (or team SA differences as measured by some other technique). High correlations would indicate that the particular representation of team process is valid in the sense that it predicts team performance. Also, if a logical or empirical team process referent is available (e.g., it may be based on an aggregate of several procedural representations from very effective teams) this can be used to assess team process by comparing the referent to each team process representation. To verify that the procedure elicits meaningful team behavior, the quantitative comparisons can be correlated with measures of team performance. It is expected that better teams should have procedural representations that more closely approximate the referent. Once the procedure and materials have been confirmed, the same referent comparison process can be used to assess teams in terms of their team process behaviors. Qualitative comparisons between the referent and the team structure may also reveal areas in which a team is particularly weak or strong. This comparison procedure can also be used to identify effects of various other factors on team behavior. For instance, the structures of two different teams or the same team performing two tasks that differ in complexity or workload could be compared.

## CONCLUSION

Recent efforts to understand and train team situation awareness have underscored the need for measures that focus on the content of SA and that extend beyond the final product of SA to measure its precursors. In this chapter we propose that knowledge elicitation methods be applied to team SA to address some of these objectives. We have taken initial steps toward this application by prescribing a set of methods for eliciting, confirming, and assessing SA, and team SA in particular. In this methodology we have described how elicitation methods can be used to elicit the content of SA and then tied these methods to other methods that can be used to confirm this content and ultimately assess team SA.

We realize, however, that this methodology only scratches the surface of team SA measurement. We hope that it provides a starting point for additional work and maybe suggests some novel measures of team SA or novel applications of elicitation methods. Further, the application of these methods to team SA has raised and will continue to raise a host of interesting theoretical, empirical, and methodological questions regarding team SA. For example, knowledge elicitation methods have traditionally targeted the individual. However, many would argue that teams and team SA in particular is more than the aggregate of the individual or of individual situation models (Prince, et al., 1994; Salas, et al., 1995). This Gestalt-like view of teams begs for a more holistic means of elicitation that is directed toward the elicitation of the team's knowledge or cognition. Can we develop or adapt elicitation methods so that they are focused more on this macro level in that they elicit knowledge from the team as a whole? If we do, will the information elicited be more (or less) predictive of team performance than an aggregate of individually elicited information.

Another issue that surfaces has to do with knowledge elicited from teams that are heterogeneous with respect to the knowledge base. Based on the definition of teams offered earlier, this role-specific orientation of individuals is more often the rule, not the exception and it poses interesting implications for measurement. For instance, similarity metrics would not necessarily be the best indicator of shared knowledge for this kind of team, nor would accuracy based on the comparison of each individual to some all-knowing referent. Initial work has been taken to examine metrics more appropriate for heterogeneous teams (Cooke, Stout, Rivera, & Salas, 1998; Cooke, et al., 2000). At the same time issues like these have raised theoretical questions about shared knowledge, cross-training, and team competencies. We are hopeful that future work on the methodological front of team SA will raise even more interesting questions relevant to team cognition in general.

## REFERENCES

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors, 37,* 85-104.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 640-645.

Blickensderfer, E., Cannon-Bowers, J. A., & Salas, E. (1997a). Training teams to self-correct: An empirical evaluation. Paper presented at the Meeting of the Society for Industrial and Organizational Psychology, St. Louis, MO (April 10-13).

Blickensderfer, E., Cannon-Bowers, J. A., & Salas, E. (1997b). Does overlap of team member knowledge predict team performance? Paper presented at the 1997 Human Factors and Ergonomics Society Annual Meeting, Albuquerque, NM (September 22-26).

Bowers, C. A., Braun, C., & Kline, P. B. (1994). Communication and team situational awareness. In R. D. Gilson, D. J. Garland, & J. M. Koonce (Eds.), *Situational awareness in complex systems* (p. 305-311). Daytona Beach, FL: Embry -Riddle Aeronautical University Press.

Brannick, M. T., Prince, A., Prince, C., & Salas, E. (1995). The measurement of team process. *Human Factors*, *37,* 641-651.

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

Carmody, D. P., Kundel, H. L., & Toto, L. C. (1984). Comparison scans while reading chest images: Taught but not practiced. *Investigative Radiology, 19*, 462-466.

Chase, W. G., & Simon, H. A. (1973a). The mind's eye in chess. In W. G. Chase (Ed.), *Cognitive skills and their acquisition* (pp. 141-189). Hillsdale, NJ: Erlbaum.

Chase, W. G., & Simon, H. A. (1973b). Perception in chess. *Cognitive Psychology, 4*, 55-81.

Cooke, N. J. (1999). Knowledge elicitation. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay, & M. T. H. Chi (Eds.), *The Handbook of Applied Cognition* (chapter 16), Chichester: John Wiley.

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies, 41,* 801-849.

Cooke, N.J. (1992a). Predicting judgment time from measures of psychological proximity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18,* 640-653.

Cooke, N. J. (1992b). Eliciting semantic relations for empirically derived networks. *International Journal of Man-Machine Studies, 37,* 721-750.

Cooke, N. M., Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*, 538-549.

Cooke, N. J., Neville, K. J., & Rowe, A. L. (1996). Procedural network representations of sequential data. *Human-Computer Interaction, 11,* 29-68.

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. (2000). Measuring team knowledge. *Human Factors, 42*, 151-173.

Cooke, N. J. & Schvaneveldt, R. W. (1988). Effects of computer programming experience on network representations of abstract programming concepts. *International Journal of Man-Machine Studies, 29,* 407-427.

Cooke, N. J., Stout, R., Rivera, K., & Salas, E. (1998). Exploring measures of team knowledge. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 215-219.

Coury, B. G, Weiland, M. Z., & Cuqlock-Knopp, V. G. (1992). Probing the mental models of system state categories with multidimensional scaling. *International Journal of Man-Machine Studies, 36,* 673-696.

Durso, F. T., & Gronlund, S. D. (1999). Situation awareness. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay, & M.

T. H. Chi (Eds.), *The Handbook of Applied Cognition* (chapter 10), Chichester: John Wiley.

Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., & Nikolic, D. & Manning, C. A. (1998). Situation awareness as a predictor of performance in en route air traffic controllers. *Air Traffic Control Quarterly.*

Endsley, M. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society's 32nd Annual Meeting*, 97-102.

Endsley, M. R. (1989). *Final report: Situation awareness in an advanced strategic mission* (NOR DOC 89-32). Hawthorne, CA: Northrop Corp.

Endsley, M. R. (1990). A methodology for the objective measure of situation awareness. In *Situational awareness in aerospace operations* (AGARD-CP-478; pp. 1/1-1/9). Neuilly-Sur-Seine, France: NATO--Advisory Group for Aerospace Research and Development.

Endsley, M. R. (1993). A survey of situation awareness requirements in air-to-air combat fighters. *The International Journal of Aviation Psychology, 3,* 157-168.

Endsley, M. R. (1995a). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37,* 32-64.

Endsley, M. R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors, 37,* 65-84.

Endsley, M. R., & Rogers, M. D. (1994). Situation awareness requirements for enroute air traffic control. *FAA Office of Aviation Medicine Reports, 94,* 27-34.

Ericsson, K. A., & Simon, H. A. (1996). *Protocol Analysis: Verbal Reports as Data* (revised edition). Cambridge, MA: MIT press.

Federico, P. (1995). Expert and novice recognition of similar situations. *Human Factors, 37,* 105-122.

Federico, P. (1997). An empirical examination of metacognitive models of situation assessment. *Human Factors, 39,* 149-157.

Fowlkes, J. E., Lane, N. E., Salas, E., Franz, T., & Oser, R. (1994). Improving the measurement of team performance: The TARGET's methodology. *Military Psychology, 6,* 47-61.

Fracker, M. L. (1991). *Measures of situation awareness: Review and future directions* (Tech. Report AL-TR-1991-0128). Wright Patterson Air Force Base, OH: Air Force Systems Command.

Fracker, M. L., & Davis, S. A. (1990). Measuring operator situation awareness and mental workload. In *Proceedings of the Fifth Mid-Central Ergonomics/Human Factors Conference*. Dayton, OH: University of Dayton.

Gaba, D. M., Howard, S. K., & Small, S. D. (1995). Situation awareness in anesthesiology. *Human Factors, 37,* 20-31.

Gillan, D. J., Breedin, S. D., & Cooke, N. M. (1992). Network and multidimensional representations of human-computer interface design experts. *International Journal of Man-Machine Studies, 36,* 587-615.

Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser, and M. J. Farr (Eds.), *The nature of expertise* (pp. xv-xxviii). Hillsdale, NJ: Erlbaum.

Goldsmith, T. E., & Davenport, D. M. (1990). Assessing structural similarity of graphs. In R. Schvaneveldt (Ed.), *Pathfinder Associative Networks: Studies in Knowledge Organization* (pp. 75-87). Norwood, NJ: Ablex.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology, 83,* 88-96.

Gray, S. H. (1990). Using protocol analysis and drawings to study mental model construction during hypertext navigation. *International Journal of Human-Computer Interaction, 2,* 359-377.

Gualtieri, J., Fowlkes, J., & Ricci, K. E. (1996). Measuring individual and team knowledge structures for use in training. *Training Research Journal, 2,* 117-141.

Harrison, B. L. (1991). Video annotation and multimedia interfaces: From theory to practice. *Proceedings of the Human Factors Society's 35th Annual Meeting*, 319-323.

Hartel, C. E. J., Smith, K., & Prince, C. (1991). Defining aircrew coordination: Searching for mishaps for meaning. Paper presented at the Sixth International Symposium on Aviation Psychology, Columbus, OH.

Hoffman, R. R., Shadbolt, N. R., Burton, A. M., & Klein, G. (1995). Eliciting knowledge from experts: A methodological analysis. *Organizational Behavior and Human Decision Processes, 62*, 129-158.

Hutchins, E.(1991). The social organization of distributed cognition. In L.B. Resnick, J. M. Levine, & S. D. Teasley (Eds), *Perspectives on Socially Shared Cognition* (pp. 283-307), Washington, DC: APA.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32*, 241-254.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge.* Hillsdale, NJ: Erlbaum.

Kempton, W. (1986). Two theories of home heat control. *Cognitive Science, 10,* 75-90.

Kirwan, B., & Ainsworth, L. K. (1992). *A Guide to Task Analysis.* London: Taylor & Francis.

Klein, G. A. (1994). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods.* NJ: Ablex.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor. *Journal of Management, 20,* 403-437.

Kraiger, K., Krause, J. R., & Hart, P. F. (1996). Construct validation of multiple measures of shared mental models. Paper presented at the Annual Convention of the American Psychological Association, Toronto, CA, August 12.

Kraiger, K., & Wenzel. L. H. (1997). Conceptual development and empirical evaluation of measures of shared mental models as indicators of team effectiveness. In M. T. Brannick, E. Salas, & C. Prince (Eds.). *Performance Assessment and Measurement: Theory, Methods, & Applications* (pp. 63-84). Mahwah, NJ: Erlbaum.

Lauwereyns, J., & d'Ydewalle, G. (1996). Knowledge acquisition in poetry criticism: The expert's eye movements as an information tool. *International Journal of Human Computer Studies, 45,* 1-18.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*,273-283.

Minionis, D. P., Zaccaro, S. J. & Perez, R. (1995). Shared mental models, team coordination, and team performance. In J. Mathieu (Chair), Mental models and team effectiveness: Three empirical tests. Symposium presented to the 10th annual conference of the Society for Industrial/Organizational Psychology, Orlando, FL

Mosier, K. L. & Chidester, T. R. (1991). Situation assessment and situation awareness in a team setting. *Proceedings of the 11th Congress of the International Ergonomics Association*, Paris, 798-800.

Orasanu, J. M. (1990). *Shared mental models and crew decision making* (CSL Tech. Rep. No. 46). Princeton, NJ: Princeton University, Cognitive Sciences Laboratory.

Owen, R. (1993). Timelines: A system for analysing time based data. Documentation, CSRI, University of Toronto.

Payne, S. J. (1991). A descriptive study of mental models. *Behaviour and Information Technology, 10,* 3-21.

Prince, C. W., Chidester, T. R., Bowers, C., & Cannon-Bowers, J. A. (1992). Aircrew coordination--Achieving teamwork in the cockpit. In R. Swezey and E. Salas (Eds.), *Teams: Their Training and Performance* (p. 329-353). Norwood, NJ: Ablex.

Prince, C., Stout, R. & Salas, R. (1994). *Team situational awareness preliminary lessons learned: Identification of skills, behaviors, design guidelines, and training strategies from interviews and research.* Technical Report, FAA and Naval Air Warfare Center, Training Systems Division.

Rasmussen, J., & Jensen, A. (1974). Mental procedures in real-life tasks: A case study of electronic troubleshooting. *Ergonomics, 17,* 293-307.

Rentsch, J. R. & Hall, R. J. (1994). Members of great teams think alike: A model of team effectiveness and schema similarity among team members. In M. M. Beyerlein & D. A. Johnson (Eds.), *Advances in interdisciplinary studies of work teams: Theories of self-managing work teams* (Vol. 1, pp. 223-262). Greenwich, CT: JAI Press.

Rouse, W. B., Cannon-Bowers, J. A., & Salas, E. (1992). The role of mental models in team performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics, 22,* 1296-1308.

Rowe, A. L., Cooke, N. J., Hall, E. P., & Halgren, T. L. (1996). Toward an on-line knowledge assessment methodology: Building on the relationship between knowing and doing. *Journal of Experimental Psychology: Applied, 2,* 31-47.

Rowe, A. L., Cooke, N. J., & Rivera, K. (1998). Assessing knowledge structures: An analysis of knowledge standards. Poster presented at the 42nd annual meeting of the Human Factors and Ergonomics Society, October 5-9, Chicago, IL.

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

Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human Factors, 37,* 123-136.

Sanderson, P. M., & Fisher, C. (1994). Exploratory sequential data analysis: Foundations. *Human-Computer Interaction, 9,* 251-317.

Sanderson, P. M., McNeese, M. D., & Zaff, B. S. (1994). Handling complex real-world data with two cognitive engineering tools: COGENT and MacSHAPA. *Behavior Research Methods, Instruments, & Computers, 26,* 117-124.

Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *International Journal of Aviation Psychology, 1,* 45-57.

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors, 37,* 5-19.

Schvaneveldt, R. W. (1990). *Pathfinder associative networks: Studies in knowledge organization.* Norwood, NJ: Ablex.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 24, pp. 249-284). New York: Academic Press.

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., Cooke, N. M., Tucker, R. G., & DeMaio, J. C. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies, 23,* 699-728.

Shepard, R. N. (1962a). Analysis of proximities: Multidimensional scaling with an unknown distance function. I *Psychometrika, 27*, 125-140.

Shepard, R. N. (1962b). Analysis of proximities: Multidimensional scaling with an unknown distance function. II *Psychometrika, 27,* 219-246.

Shrestha, L. B., Prince, C., Baker, D. P., & Salas, E. (1995). Understanding situation awareness: Concepts, methods, and training. In W. B. Rouse (Ed.), *Human/technology interaction in complex systems, vol. 7* (pp. 45-83). San Francisco: JAI Press.

Stout, R. J. (1995). Planning effects on communication strategies: A shared mental models perspective. Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting, Santa Monica, CA, 1278-1282.

Stout, R., Cannon-Bowers, J. A., & Salas, E. (1996). The role of shared mental models in developing team situation awareness: Implications for training. *Training Research Journal, 2,* 85-116.

Taylor, R. M. (1989). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Proceedings of the NATO*

*AGARD Conference on Situational Awareness in Aerospace Operations* (AGARD-CP-478). Springfield, VA: National Technical Information Service.

Taylor, R. M. & Selcon, S. J. (1994). Situation in mind: Theory, application, and measurement of situational awareness. In R. D. Gilson, D. J. Garland, & J. M. Koonce (Eds.), *Situational awareness in complex systems* (p. 69-77). Daytona Beach, FL: Embry -Riddle Aeronautical University Press.

vanSomeren, M. W., Barnard, Y.F., & Sandberg, J.A.C. (1994). *The Think Aloud Method: A Practical Guide to Modeling Cognitive Processes*. London: Academic Press.

Vidulich, M. A. (1989). The use of judgment matrices in subjective workload assessment: The subjective workload dominance (SWORD) technique. *In Proceedings of the Human Factors Society 33rd Annual Meeting* (pp. 1406-1410). Santa Monica, CA: Human Factors and Ergonomics Society.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24,* 535-585.

Walsh, J. P., Henderson, C. M., & Deighton, J. (1988). Negotiated belief structure and decision performance: An empirical investigation. *Organizational Behavior and Human Decision Processes, 42,* 194-216.

Wickens, C. D., Bellenkes, A. H., & Kramer, A. F. (1995). Visual scanning expert pilots: The role of attentional flexibility and mental model development. University of Illinois at Urbana-Champaign: Human Perception and Performance Technical Report UIUC-BI-HPP-95-05.

Individual SA

- Situation model - individual's understanding of situation (i.e., the product of SA)
- Mental models - knowledge (a precursor product)
- Cues - information in the environment (a precursor product)
- Situation assessment process (a precursor process)

Team SA

- Team situation model - team's understanding of situation (i.e., the product of team SA)
- Situation model of each team member (a precursor product)
- Shared mental models (a precursor product)
- Team process behaviors (a precursor process)

**Table 1.** Targets of team SA measurement.

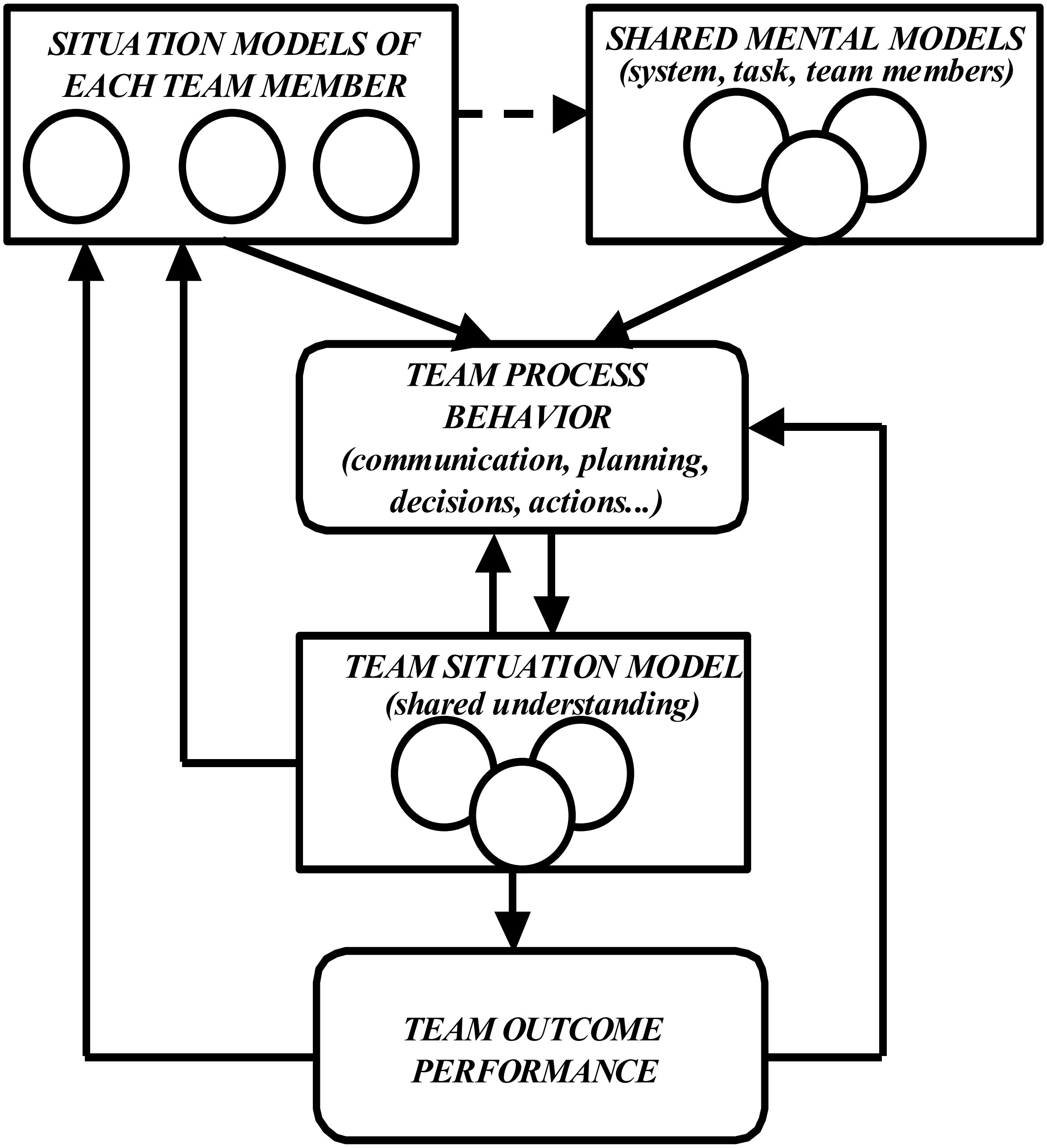
| Target | Elicitation Methods | Elicitation Products |
|---|---|---|
| **1. Situation Models** | | |
| **Individual** | Structured interviews | Raw or categorized interview responses |
| **Team** | Intrateam comparison & aggregation of structured interview | Intrateam summaries of structured interviews |
| **2. Mental Models** | | |
| **Individual** | Structured interviews | Raw or categorized interview responses |
| **System:** | component-function diagrams | |
| **Task:** | goal decomposition | |
| **Team Members :** | information flow diagram<br>role reversal | |
| | Conceptual methods | Representations of individual knowledge structures |
| **Shared** | Intrateam comparison & aggregation of knowledge structures | Team knowledge structure |
| **3. Cues & Situation Assessment (Individual)** | Observations | Hypothetical cues |
| | Conceptual methods | Hypothetical cues |
| | Process tracing | Cues, cue categories, processes underlying situation assessment |
| **4. Team Process Behaviors (Team)** | Process tracing | Behavioral protocol, frequencies, procedural representation |

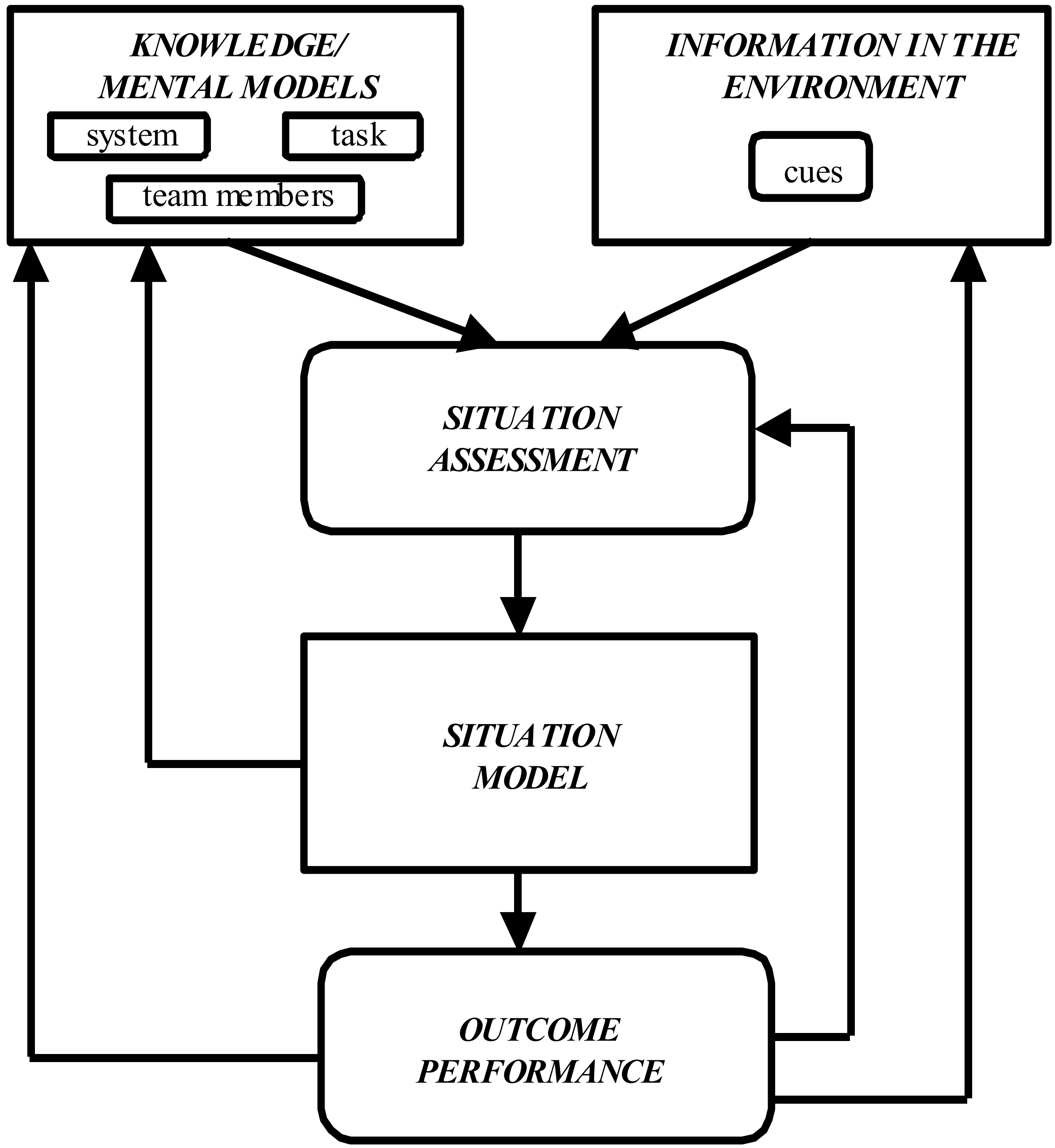
**Table 2a.** Targets for team SA measurement and prescribed elicitation methods for this purpose.

| Target | Assessment Methods | Assessment Metrics |
|---|---|---|
| **1. Situation Models** | | |
| **Individual** | Queries | Query accuracy |
| **Team** | Team-oriented queries and intrateam comparison of query responses | Intrateam query accuracy & intrateam similarity |
| **2. Mental Models** | | |
| **Individual** | Compare knowledge structure to individual referent | Accuracy based on similarity to referent |
| **Shared** | Compare team knowledge to team referent and intrateam comparison of knowledge structures | Team accuracy based on similarity to team referent; Intrateam similarity |
| **3. Cues & Situation Assessment** | | |
| **(Individual)** | Tracking eye movements | Presence or absence of eye movements associated with a cue |
| | Policy capturing | Regression weights associated with cues |
| | Experimental manipulation | Differences in task performance with manipulation |
| **4. Team Process Behaviors** | | |
| **(Team)** | Compare team process representation to team referent | Team process effectiveness based on similarity to team referent |

**Table 2b.** Targets for team SA measurement and prescribed assessment methods for this purpose.

## FIGURE LIST


SITUATION MODELS OF EACH TEAM MEMBER
SHARED MENTAL MODELS (system, task, team members)
TEAM PROCESS BEHAVIOR (communication, planning, decisions, actions...)
TEAM SITUATION MODEL (shared understanding)
TEAM OUTCOME PERFORMANCE


KNOWLEDGE/
MENTAL MODELS
system
task
team members
INFORMATION IN THE
ENVIRONMENT
cues
SITUATION
ASSESSMENT
SITUATION
MODEL
OUTCOME
PERFORMANCE

# AUTHOR BIOGRAPHIES

**Nancy J. Cooke**
Department of Psychology
New Mexico State University
Box 30001, MS 3452
Las Cruces, NM 88003

Phone: 505-646-1630
Fax: 505-646-6212
Email: cooke@crl.nmsu.edu

Nancy J. Cooke is a Professor of Psychology at New Mexico University, a Carnegie I Hispanic serving minority institution. She received her Ph.D. in cognitive psychology in 1987 from New Mexico State University, served on the Psychology faculty at Rice University from 1987-1992, and returned to New Mexico State in 1992. She teaches courses in cognitive and engineering psychology and supervises graduate and undergraduate research in the Cognitive Engineering Laboratory and the CERTT Laboratory for Cognitive Engineering Research on Team Tasks in which measures of team cognition are developed, applied, and evaluated to ultimately enhance scientific understanding of team cognition. Dr. Cooke's research interests include the study of knowledge and its application to the development of cognitive and knowledge engineering methodologies, as well as to expertise, intelligent tutors, human-computer interfaces, and the assessment of individual and team cognition. Dr. Cooke has experience conceptualizing, designing, and directing research in a variety of applied settings, including NASA-JSC, XSOFT, Sperry Corporation, California Peace Officer Standards and Training Commission, U.S. Department of Labor and her most recent research has been funded by AFOSR, Naval Air Warfare Center Training Systems Division, and ONR. She is also on the editorial boards of *Human Factors, International Journal of Human-Computer Studies* and *Journal of Experimental Psychology: Applied.*

**Rene'e J. Stout**
AlignMark
Maitland, FL

Fax: 407-659-3580
Email: rstout@alignmark.com

Renée J. Stout is the Performance Development Manager at AlignMark, a Human Resources Consulting firm located in Maitland, FL. Prior to her current position, she was a Research Psychologist at the Naval Air Warfare Center in Orlando, FL, for more than ten years. She received her Ph.D. from the University of Central Florida in Human Factors Psychology in August, 1994. She has worked in the area of training design and development and human performance measurement for more than 13 years. She has over 100 presentations and publications in the areas of training, team training, and cognitive influences on performance and was twice the recipient of the Human Factors and Ergonomics Society Training (HFES) Technical Group award for best student paper. She was also a past Program Chair for the Training Technical Group of HFES.

**Eduardo Salas**
Department of Psychology
University of Central Florida
Box 161390
Orlando, FL 32816-1390

Phone: 407-823-2552
Fax: 407-823-5862
Email: esalas@pegasus.cc.ucf.edu

SHOULD HAVE THIS ALREADY