2010

Integrating Situation Awareness Assessment Into Test and Evaluation

Cheryl A. Bolstad
SA Technologies, Marietta, Georgia

Haydee Maria Cuevas
Embry-Riddle Aeronautical University, cuevash1@erau.edu

ITEA Journal 2010; 31: 240–246

Integrating Situation Awareness Assessment Into Test and Evaluation

Cheryl A. Bolstad, Ph.D. and Haydee M. Cuevas, Ph.D.
SA Technologies, Marietta, Georgia

To guarantee the success of network-centric operations, warfighters need the ability to extract and share critical task-relevant information to develop and maintain the situation awareness that is so critical for effective team performance. As such, the design of emerging technologies and systems must adopt a "user-centric" approach, with consideration for human information processing capabilities and limitations. In turn, to ensure that these technologies and systems are meeting their design objectives, test and evaluation must similarly be expanded to include metrics that assess how well system features and functions are supporting critical human cognitive processes such as situation awareness and decision-making. In this article, we address this issue, focusing specifically on situation awareness. We discuss how situation awareness assessment, at both the individual and team level, can be integrated into test and evaluation. We also cite examples from our own research to demonstrate the diagnosticity afforded by situation awareness assessment.

Key words: Decision-making, diagnostics, human cognition, information technology, network centric warfare, team performance.

Network centric warfare promises to provide revolutionary command, control, and communications capabilities. With this increased network-centricity, the state of current military operations is shifting from traditional large command and control centers to small groups working together in a distributed manner through the use of information technology. Although advances in information technology are enabling this drive toward network-centricity through the development of networked databases, greater bandwidths, and more sophisticated collaboration tools, the deciding factor is how human operators will be able to work collaboratively to capitalize on this enormous volume of available information (Lawlor 2005).

To guarantee the success of network-centric operations, warfighters need the ability to extract and share critical task-relevant information to develop and maintain the situation awareness (SA) that is so critical for effective team performance. As such, the design of emerging technologies and systems must adopt a *user-centric* approach, with consideration for human information processing capabilities and limitations. In turn, to ensure that these technologies and systems are meeting their design objectives, test and evaluation must similarly be expanded to include metrics that assess how well system features and functions are supporting critical human cognitive processes such as SA and decision-making. In this article, we address this issue, focusing specifically on SA. We begin with a brief overview of SA, defining this construct at both the individual and team level. We then discuss how SA assessment can be integrated into test and evaluation, citing examples from our own research.

SA defined

Although several different definitions of situation awareness have been put forth in the literature (Fracker 1991; Sarter and Woods 1991; Smith and Hancock 1995), in this article, we focus on Endsley's (1995b) theoretical model of SA, which defines this complex cognitive construct as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley 1995b, 36). As implied by this definition, SA involves being aware of what is happening around you to understand how information, events, and your own actions will affect your goals and objectives, both now and in the near future. Endsley's definition highlights three levels of SA: perception, comprehension, and projection.

Perception (Level 1 SA) involves an active process whereby individuals extract significant cues from their environment, selectively directing attention to important information, while disregarding nonrelevant items. *Comprehension* (Level 2 SA) involves integrating this information in working memory to understand how the information will influence the individual's goals and objectives. *Projection* (Level 3 SA) involves extrapolating this information forward in time to determine how it will affect future states of the operating environment. Consideration of these three levels of SA is useful for understanding the types of difficulties human operators face while performing their tasks and also for determining how best to mitigate these challenges.

At the team level, SA can be viewed in terms of both team SA and shared SA. *Team SA* can be defined as "the degree to which every team member possesses the SA required for his or her responsibilities" (Endsley 1995b, 39). Thus, to ensure successful performance, each team member needs to have superior SA on those factors that are relevant for his or her job. In contrast, *shared SA* can be defined as "the degree to which team members possess the same SA on shared SA requirements" (Endsley and Jones 2001, 48). A major part of teamwork involves understanding the SA requirements that are relevant across multiple team members. Successful team performance, therefore, is influenced by the degree to which team members share a common understanding of what is happening on these shared SA elements. In other words, team members must be able to access and similarly interpret important information on the shared SA requirements that are relevant across their different positions.

Role of SA in human performance

SA represents one of the most challenging aspects of human performance. In particular, in most complex tasks, effective decision-making largely depends upon the degree to which individuals have developed a good understanding of the situation, namely, their SA. SA is especially crucial in domains where information flow can be quite high and poor decisions may lead to serious consequences (e.g., piloting an airplane, functioning as a soldier, treating critically ill or injured patients). Indeed, SA has been recognized as a critical, yet often elusive, foundation for successful decision-making across a broad range of complex and dynamic systems, including aviation and air traffic control (Nullmeyer et al. 2005), emergency response and military command and control operations (Blandford and Wong 2004; Gorman, Cooke, and Winner 2006), and offshore oil and nuclear power plant management (Flin and O'Connor 2001). Lacking SA or having inadequate SA has been consistently identified as one of the primary factors in accidents attributed to human error (Hartel, Smith, and Prince 1991; Merket, Bergondy, and Cuevas-Mesa 1997; Nullmeyer et al. 2005). Yet, developing and maintaining SA imposes high cognitive demands upon human operators in terms of time, attention, and effort. Fortunately, the cognitive load associated with achieving high levels of SA can be mitigated through SA-oriented system design (see Endsley, Bolte, and Jones 2003) and SA-oriented training programs (Strater and Bolstad 2009). Hence, test and evaluation plays a major role in ensuring that these systems achieve their design objectives, a topic we turn to next.

SA assessment

An SA-oriented approach to test and evaluation goes beyond simply assessing a system's functional capabilities to also include how well the system's design supports human operators' critical cognitive processes underlying SA and decision-making. At the team-level, this also includes evaluating the system's effectiveness in supporting the team's ability to assess and track coordination, communication, collaboration, and information-sharing activities. In general, methodologies to assess SA vary in terms of direct measurement (e.g., objective real-time probes or subjective questionnaires assessing perceived SA) or indirect methods (e.g., process indices, trained observer ratings) that infer SA based on operator physiological state, behavior, or performance. Direct measures are typically considered to be "product-oriented" in that these techniques assess an SA outcome; indirect measures are considered to be "process-oriented," focusing on the underlying processes or mechanisms required to achieve SA (Graham and Matthews 2000). Selecting which methodology to use depends upon the researcher's objectives and what data collection facilities or setup is available. Examples of each of these SA measurement approaches will be further described next.

Process indices

Process indices, such as psycho-physiological measures, examine how individuals process information in their environment (Wilson 2000). Such measures include electroencephalography (EEG), event-related potentials (ERP), event-related desynchronization (ERD), heart rate variability (HRV), electrodermal activity (EDA), eye blinks, and eye tracking. Tracking eye movements, in particular, is one of the more common psycho-physiological approaches for providing insight into perception and comprehension. Eye-tracking devices can be used to monitor where operators are directing their attention, and thereby, determine whether the saliency of important cues is sufficient or if nonessential cues are drawing away the operator's attention. Analyzing communications can also serve as process indices of operator SA. For example, verbalizations made by operators during a task can be analyzed to determine how well information is being acquired from a system designed to support this task.

Process indices are advantageous in that these offer objective assessment of operator SA and provide an indication of information access and utilization. However, process indices create large amounts of data to analyze and are difficult to implement in the real-world environment (e.g., eye-tracking devices, if head-mounted, can be cumbersome and intrusive). Further, process indices do not directly assess SA but rather can only be used to infer SA. In other words, these measures do not indicate what is actually done with the information acquired (processing) or whether the information is registered correctly or what is retained in memory. Instead, these measures simply indicate that the operator looked at the information. Given these limitations, process indices are more suitable for investigating specific research questions of information acquisition and for examining the processes underlying SA (e.g., perception, attention) rather than the final product.

Subjective measures

Subjective measures directly assess SA by asking individuals (or experienced observers) to rate their SA on an anchored scale (for a detailed review, see Jones 2000). These ratings can be collected during task performance or following task completion. Subjective measures of SA are attractive in that they are relatively straightforward, inexpensive, and easy to administer. However, several important limitations should be noted. Individuals making subjective assessments of their own SA are often unaware of information they do not know. Further, self-ratings may be tainted by performance outcomes. Subjective measures also tend to be global in nature and, as such, do not fully exploit the multivariate nature of SA to provide the detailed diagnostics available with objective measures. Nevertheless, self-ratings may be useful in that they can provide an assessment of operators' degree of confidence in their SA.

Subjective estimates of an individual's SA may also be made by experienced observers (e.g., supervisors, trained external experts). These observer ratings may be somewhat superior to self-ratings of SA because more information about the true state of the environment is usually available to the observer than to the operator, who may be focused on performing the task (i.e., trained observers may have more complete knowledge of the situation). However, observers have only limited knowledge about the operator's concept of the situation and cannot have complete insight into the mental state of the individual being evaluated. Thus, observers are forced to rely more on operators' observable actions and verbalizations in order to infer their level of SA. In this case, such actions and verbalizations are best assessed using performance and behavioral measures of SA, as described next.

Performance and behavioral measures

Performance measures infer SA from the end result (i.e., task performance outcomes) based on the assumption that better performance indicates better SA. Common performance metrics include quantity of output or productivity level, time to perform the task or respond to an event, and the accuracy of the response or, conversely, the number of errors committed. The main advantage of performance measures is that these can be collected objectively and without disrupting task performance. However, although evidence exists to suggest a positive relation between SA and performance, this connection is probabilistic and not always direct and unequivocal (Endsley 1995b). In other words, good SA does not always lead to good performance, and poor SA does not always lead to poor performance (Endsley 1990). Thus, performance measures should be used in conjunction with others measures of SA that directly assess this construct.

Behavioral measures also infer SA from the actions that individuals choose to take, based on the assumption that good actions will follow from good SA and vice versa. Behavioral measures rely primarily on observer ratings and are thus somewhat subjective in nature. To address this limitation, observers can be asked to evaluate the degree to which individuals are carrying out actions and exhibiting behaviors that would be expected to promote the achievement of higher levels of SA. This approach removes some of the subjectivity associated with making judgments about an individual's internal state of knowledge by allowing them to make judgments about SA indicators that are more readily observable.

Objective measures

Objective measures directly assess SA by comparing an individual's perceptions of the situation or environment with some "ground truth" reality. Specifically, objective measures can be used to collect data from operators' perceptions of the situation and compare this with what is actually happening at a given moment in time. Thus, this type of assessment provides a direct measure of SA and does not require operators or experimenters to make judgments about situational knowledge on the basis of incomplete information. Objective measures can be gathered in one of three ways: during an interruption in task performance (e.g., queries), real time as the task is completed (e.g., probes), or posttest following completion of the task.

One common approach to directly and objectively measure SA is the Situation Awareness Global Assessment Technique (SAGAT) (Endsley 1995a). SAGAT utilizes a concurrent memory probe technique that presents queries related to the current task environment. Administration of the SAGAT involves freezing a simulation exercise at randomly selected times and hiding task information sources (e.g., blanking visual displays) while individuals quickly answer randomly ordered questions about their current perceptions of the situation. These responses are then compared with "ground truth" (i.e., actual data on the real situation) to assess the accuracy of the individuals' SA. However, because it involves interrupting task performance, SAGAT is better suited for assessing SA in simulation exercises and may not be practical for real-time measurement of SA.

For settings in which disruptions to task performance are not practical or desirable, real-time probes (e.g., open-ended questions embedded as verbal communications during the task) can be administered to naturally and unobtrusively assess operator SA (Jones and Endsley 2000). Real-time probes are similar to SAGAT in that they query operators on their knowledge of key task-relevant information in the environment; however, this methodology differs from the SAGAT in that task performance is not disrupted (i.e., the simulation or task is not stopped) but rather the queries are incorporated as a natural part of the task.

Modeling SA

SA modeling approaches can be used to objectively predict SA based on readily observable verbal and nonverbal communications. Specifically, team communications (particularly verbal communications) support the knowledge building and information processing that lead to SA construction (Endsley and Jones 2001). Thus, since SA may be distributed via communication, computational linguistics and machine learning techniques can be combined with natural language analytical techniques (e.g., Latent Semantic Analysis) to create models that draw on the verbal expressions of the team to predict SA and task performance (Bolstad et al. 2005b, 2007). For example, the Automated Communication and Situation Awareness (ACASA) tool offers near real-time, nonintrusive, quantitative assessment of SA by analyzing communication exchanges among team members (Foltz et al. 2008). Since the communication data are collected using either Automatic Speech Recognition (ASR) software or transcriptions of speech recordings, this methodology does not interrupt activities or affect performance. Thus, SA modeling approaches, such as the ACASA tool, are appropriate for use in both simulations and real-world environments. Further, this methodology can provide diagnostic information regarding current SA. For example, when coupled with ASR software, the ACASA tool can be used to quickly identify whether or not immediate action needs to be taken to address poor SA among team members.

Although evidence exists to support the utility of communication analysis for predicting team SA (Foltz et al. 2008), time constraints and technological limitations (e.g., cost and availability of speech recording systems and speech-to-text translation software) may make this approach more time consuming in terms of up-front investment. In addition, the models generated using this approach are domain- and task-specific; thus, unique models must be created for each environment or application. Last, this measure is only effective for measuring SA in a team environment and would not be suitable for situations in which a single operator is being evaluated.

Applying SA assessment to teams

Not surprising, assessing team and shared SA is more complex than assessing SA at the individual level. Some methodologies are inherently more readily applicable for team-level assessment. For example, the ACASA tool described earlier is specifically designed to be applied in a team context; thus it can be used to evaluate information flow during task performance in terms of how well team members are sharing the SA information requirements necessary for building and maintaining both team and shared SA. Similarly, behavioral measures can be used to support assessment of the types of overt team behaviors and communications that are indicative of SA.

Comparison of individual responses to objective measures of SA (e.g., SAGAT queries or real-time probes) across different team members can be used to ascertain the degree to which they have developed a common and accurate understanding of the situation or task environment (i.e., shared SA). Thus, this approach can provide the degree of diagnosticity needed to fully evaluate team performance. The simplest analysis involves comparing performance between two team members. For example, when analyzing two team members' responses to a SAGAT query, one of four possible outcomes can occur: both individuals are correct; one individual is correct and the other is incorrect; both individuals are incorrect and they have the same response; or both individuals are incorrect but they have different responses (Endsley and Jones 2001). The latter three outcomes highlight different problems with the team members' shared SA, which in turn can provide insights on how to address this potential breakdown in team performance.

Lessons learned in SA assessment

Our work on assessing SA in team operations has demonstrated that using multiple metrics provides the greatest utility in terms of understanding how and why teams perform. For example, in a brigade-level simulated military exercise, we utilized the SAGAT methodology to evaluate a possible new unit formation (Bolstad and Endsley, In press). While the overall exercise was deemed a success, analysis of our SA assessment results indicated that placing the Deputy Brigade Commander away from the Commander hindered his ability to develop the same level of SA as the Commander. In another military exercise, we evaluated using cross-training as a method to improve team SA and performance (Bolstad et al. 2005a, 2005b). In addition to administering an objective measure of SA (i.e., SAGAT queries), we also included a subjective measure of team communication that specifically asked participants to rank order other team members based on their frequency of communication with them during the scenario; this measure was used to calculate social network distance, that is, the frequency with which team members communicated with each other. While results showed that cross-training, particularly in a leadership role, did lead to improved SA, analysis of the communication data also provided some insights on potential factors influencing team SA and performance. Specifically, physical distance (i.e., whether participants were co-located or distributed during the scenario) was found to be a significant predictor of both shared SA and social network distance. This finding supports the view that direct information exchange may be used as an input for building a team member's individual SA (Endsley and Jones 2001; Milham, Barnett, and Oser 2000).

One important lesson learned from these research studies is that increasing the sensitivity and diagnosticity of test and evaluation involves adopting a multi-faceted approach to assessment. Rather than rely on a single approach or metric, valid and reliable assessment should utilize a battery of distinct yet related measures that complement each other; this approach capitalizes on the strengths of each measure while minimizing the limitations inherent in each. Combining multiple measures together can provide valuable information with regard to factors influencing team SA, decision-making, and performance, such as the effect of team organization, distribution, and communication patterns.

Conclusion

Assessment of SA provides a degree of diagnosticity that is especially useful in the test and evaluation of new technologies and systems. SA assessment can be used to identify the source of the problem as well as to establish a baseline for comparison of the effects of different design concepts. More specifically, integrating SA assessment into test and evaluation can allow researchers to determine if a new technology or system is helping or hindering human operators' ability to perceive critical information (Level 1 SA), comprehend the relevance of this information to their task (Level 2 SA), and use this information to predict what will happen next (Level 3 SA); as well as to evaluate how these effects on operator SA influence decision-making and, ultimately, safety and performance. At the team-level, SA assessment can be used to determine the degree to which information is being exchanged among team members to support both team and shared SA.

In both cases, determining the best SA measures for inclusion in Mission-Based Test and Evaluation events depends upon multiple factors, such as the study's objectives, team size, other variables being assessed, and the ability to integrate the selected measures into the experimental test plan. However, whenever possible, a multi-faceted approach to SA assessment is desirable to ensure a higher level of diagnosticity in the overall assessment. ❑

DR. CHERYL A. BOLSTAD is a principle research associate at SA Technologies, headquartered in Marietta, Georgia. She received her doctor of philosophy degree in psychology, specializing in cognition and aging, from North Carolina State University (Raleigh, North Carolina). Dr. Bolstad has 20 years of experience as a human factors engineer and has worked on a wide variety of projects including team performance analysis and measurement, training program design and evaluation, collaboration tool design, and cognitive readiness assessment. More recently, she has worked extensively in SA research including user interface design, training, and measurement. E-mail: cheryl@satechnologies.com

DR. HAYDEE M. CUEVAS is a research associate II at SA Technologies, headquartered in Marietta, Georgia. She received her doctor of philosophy degree in applied experimental and human factors psychology from the University of

Central Florida (Orlando, Florida). Dr. Cuevas has over 10 years of experience as a human factors researcher and has worked on projects funded by the National Science Foundation, Air Force Office of Scientific Research, Army Research Laboratory, Office of Naval Research, and Office of the Secretary of Defense. Her recent research has primarily focused on supporting human-automation team performance in complex operational environments. E-mail: haydee.cuevas@satechnologies.com

References

Blandford, A., and W. Wong. 2004. Situation awareness in emergency medical dispatch. *International Journal of Human–Computer Studies.* 61: 421–452.

Bolstad, C. A., and M. Endsley. In press. Measuring shared and team situation awareness in the U.S. Army's Future Objective Force. *Military Psychology.*

Bolstad, C. A., H. M. Cuevas, A. M. Costello, and J. Rousey. 2005a. "Improving situation awareness through cross-training." In *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society*, September 26–30, Orlando, Florida, 2159–2163. Santa Monica, CA: Human Factors and Ergonomic Society.

Bolstad, C. A., H. M. Cuevas, C. Gonzalez, and M. Schneider. 2005b. Modeling shared situation awareness. In *Proceedings of the 14th Conference on Behavior Representation in Modeling and Simulation*, May 16–19, Los Angeles, California.

Bolstad, C. A., P. Foltz, M. Franzke, H. M. Cuevas, M. Rosenstein, and A. M. Costello. 2007. Predicting situation awareness from team communications. In *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting*, October 1–5, Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R. 1990. Predictive utility of an objective measure of situation awareness. In *Proceedings of the Human Factors Society 34th Annual Meeting*, Santa Monica, CA, 41–45. Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R. 1995a. Measurement of situation awareness in dynamic systems. *Human Factors.* 37 (1): 65–84.

Endsley, M. R. 1995b. Toward a theory of situation awareness in dynamic systems. *Human Factors.* 37 (1): 32–64.

Endsley, M. R., and W. M. Jones. 2001. A model of inter- and intrateam situation awareness: Implications for design, training and measurement. In *New trends in cooperative activities: Understanding system dynamics in complex environments.* Edited by M. McNeese, E. Salas, and M. R. Endsley, 46–67. Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R., B. Bolte, and D. G. Jones. 2003. *Designing for situation awareness: An approach to human-centered design.* London: Taylor & Francis.

Flin, R., and P. O'Connor. 2001. Applying crew resource management in offshore oil platforms. In *Improving teamwork in organization: Applications of resource management training.* Edited by E. Salas, C. A. Bowers, and E. Edens, 217–233. Hillsdale, NJ: Erlbaum.

Foltz, P. W., C. A. Bolstad, H. M. Cuevas, M. Franzke, M. Rosenstein, and A. M. Costello. 2008. Measuring situation awareness through automated communication analysis. In *Macrocognition in teams.* Edited by M. P. Letsky, N. W. Warner, S. M. Fiore, and C. A. P. Smith, 259–275. Aldershot, England: Ashgate.

Fracker, M. L. 1991. *Measures of situation awareness: Review and future directions* (Report No. AL-TR-1991-0128). Wright-Patterson Air Force Base, OH: Armstrong Laboratories.

Gorman, J. C., N. J. Cooke, and J. L. Winner. 2006. Measuring team situation awareness in decentralized command and control environments. *Ergonomics.* 49 (12–13): 1312–1325.

Graham, S. E., and M. D. Matthews. 2000. Modeling and measuring situation awareness. In *Workshop on assessing and measuring training performance effectiveness* (Tech. Rep. 1116). Edited by J. H. Hiller, and R. L. Wampler, 14–24. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hartel, C. E. J., K. Smith, and C. Prince. 1991. Defining aircrew coordination: Searching mishaps for meaning. Paper presented at the 6th International Symposium on Aviation Psychology, April 29–May 2. Columbus, Ohio.

Jones, D. G. 2000. Subjective measures of situation awareness. In *Situation awareness analysis and measurement.* Edited by M. R. Endsley and D. J. Garland, 113–128. Mahwah, NJ: Lawrence Erlbaum.

Jones, D. G., and M. R. Endsley. 2000. Examining the validity of real-time probes as a metric of situation awareness. In *Proceedings of the 14th Triennial Congress of the International Ergonomics Association and the 44th Annual Meeting of the Human Factors and Ergonomics Society.* July 30–August 4, San Diego, CA. Santa Monica, CA: Human Factors and Ergonomics Society.

Lawlor, M. 2005. Researchers investigate cognitive collaboration. *Signal.* 59 (9): 30–34.

Merket, D. C., M. Bergondy, and H. Cuevas-Mesa. 1997. Making sense out of teamwork errors in complex environments. Paper presented at the 18th Annual Industrial/Organizational–Organizational Behavior

Graduate Student Conference, March. Roanoke, Virginia.

Milham, L. M., J. S. Barnett, and R. L. Oser. 2000. Application of an event-based situation awareness methodology: Measuring situation awareness in an operational context. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society*, July 20–August 4, San Diego, California, 2, 432–426. Santa Monica, CA: Human Factors and Ergonomics Society.

Nullmeyer, R. T., D. Stella, G. A. Montijo, and S. W. Harden. 2005. Human factors in Air Force flight mishaps: Implications for change. In *Proceedings of the 27th Annual Interservice/Industry Training, Simulation, and Education Conference* (Paper no. 2260), Arlington, VA: National Training Systems Association.

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

Smith, K., and P. A. Hancock. 1995. Situation awareness is adaptive, externally directed consciousness. *Human Factors*. 37 (1): 137–148.

Strater, L. D., and C. A. Bolstad. 2009. Situation awareness in simulations. In *Human factors in simulation and training*. Edited by D. A. Vincenzi, J. A. Wise, M. Mustapha, and P. A. Hancock, 129–148. New York, NY: CRC Press, Taylor and Francis Group.

Wilson, G. F. 2000. Strategies for psychophysiological assessment of situation awareness. In *Situation awareness analysis and measurement*. Edited by M. R. Endsley and D. J. Garland, 175–188. Mahwah, NJ: Lawrence Erlbaum Associates.