

# Measuring Situation Awareness in Complex Systems: Comparison of measures study

Paul M. Salmon\*, Neville A. Stanton, Guy H. Walker, Daniel Jenkins, Darshna Ladva,

Laura Rafferty & Mark Young

*Human Factors Integration Defence Technology Centre (HFI-DTC)  
Brunel University, BIT Lab, School of Engineering and Design, Uxbridge,  
Middlesex, UK, UB8 3PH*

\*corresponding author, [paul.salmon@brunel.ac.uk](mailto:paul.salmon@brunel.ac.uk)

## Abstract

Situation Awareness (SA) is a distinct critical commodity for teams working in complex industrial systems and its measurement is a key provision in system, procedural and training design efforts. This article describes a study that was undertaken in order to compare three different SA measures (a freeze probe recall approach, a post trial subjective rating approach and a critical incident interview technique) when used to assess participant SA during a military planning task. The results indicate that only the freeze probe recall method produced a statistically significant correlation with performance on the planning task and also that there was no significant correlation between the three methods, which suggests that they were effectively measuring different things during the trials. In conclusion, the findings, whilst raising doubts over the validity of post trial subjective rating and interview-based approaches, offer validation evidence for the use of freeze probe recall approaches to measure SA. The findings are subsequently discussed with regard to their implications for the future measurement of SA in complex collaborative systems.

## **Relevance to Industry**

Situation Awareness is a critical commodity for teams working in industrial systems and designers and analysts alike require reliable and valid methods for assessing the impact of new systems, interfaces, training programs and procedures on the level of situation awareness held by teams. This article presents a review and comparison of situation awareness measurement approaches for use in complex industrial systems and provides recommendations on the types of methods to use during situation awareness assessments.

*Keywords:* Situation Awareness, Situation Awareness Measurement, Command and Control

## **Introduction**

Situation Awareness (SA) continues to receive a considerable amount of attention from the human factors and ergonomics community (e.g. Riley, Endsley, Bolstad & Cuevas, 2006, Salmon, Stanton, Walker & Green, 2006; Stanton et al, 2006 etc). SA is the term given to the awareness that an individual has of a situation, an operator's dynamic understanding of 'what is going on' (Endsley, 1995a). Of the various SA models presented in the literature, Endsley's (1995a) information processing based three-level model is the most popular, describing SA is a *product* comprising three hierarchical levels: *level 1*, the perception of task relevant elements in the environment; *level 2*, the comprehension of their meaning in relation to task goals; and *level 3*, the projection of their future states. Endsley (1995a) formally 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" (Endsley, 1995a, p.36).

The level of SA that teams possess is now recognized as a critical factor in system design and assessment (Shu and Furuta, 2005). It follows then that system designers and also training and procedure designers need to be able to accurately describe and measure individual and team SA, as do researchers and academic wishing to evaluate SA in collaborative environments. SA measurement is therefore critical, not only to the advancement of SA-related theory, but also to system, procedure and training program design and evaluation efforts. Researchers need valid and reliable methods of assessing operator SA in order to test and advance SA theory, whilst system, procedure and training program designers need ways of assuring that SA is improved and not degraded by a new system, interface, procedure or training program.

As with most other human factors constructs (e.g. mental workload, human error, teamwork etc), there are a plethora of different approaches available to practitioners wishing to assess SA, however, there is great debate between researchers over the extent to which these measures are in fact measuring SA (and not some other psychological construct) and also which of these approaches is the most appropriate for assessing SA, particularly relating to the assessment of SA during collaborative endeavour (e.g. Gorman, Cooke & Winner, 2006; Patrick, James, Ahmed & Halliday, 2006; Salmon, Stanton, Walker and Green, 2006).

The purpose of this study was to examine the measurement of SA during a military planning task. Specifically, we wished to compare three different methods of measuring SA in order to identify the most suitable approach to use in future SA-related studies in complex military environments. Of interest to us in this research was not only the ability of the different methods to accurately and validly measure SA, but

also the ways in which the different methods view and represent SA and the extent to which they were measuring SA and not some other construct.

### **What do we want to measure anyway? A note on situation awareness theory**

It is worth pointing out at the onset that the current contention surrounding the construct of SA makes measuring it a problematic task. The SA literature is somewhat disparate, and many academics disagree on what SA actually is. For example, some view SA as knowledge in working memory (e.g. Bell & Lyon, 2000), some view it as a cognitive product of information-processing (e.g. Endsley, 1995a) and some view it as externally directed consciousness (e.g. Smith and Hancock, 1995). Further, much debate is present in the literature over whether SA refers to the *process* of gaining awareness (e.g. Fracker, 1991), the *product* of awareness (e.g. Endsley, 1995a), or a combination of the two (e.g. Smith and Hancock, 1995). To complicate matters further, SA models also differ in terms of their underpinning psychological approach. For example, the three level model (Endsley, 1995a) is a cognitive theory that uses an information processing approach, Smith and Hancock's (1995) model is an ecological approach underpinned by the perceptual cycle model (Niesser, 1976) approach, and Bedny and Meister's (1999) model uses an activity theory model to describe SA. Finally, others question the extent to which SA represents a unique psychological construct in its own right (e.g. Bell & Lyon, 2000; Moray, 2004) rather than merely being a catch all term encompassing various elements of human cognition (e.g. perception, working memory etc).

The contention surrounding the construct therefore makes the assessment of SA a complex proposition and the lack of a universally accepted model of SA (although

Endsley's three level model does come close) makes it difficult to select a measure that is congruent with the way in which one views the construct. It is notable that most of the models presented in the literature appear to have elements of truth in them (Stanton, Chambers & Piggott, 2001) and so the selection of an appropriate SA measure is difficult, since most measures are developed in line with a specific model. For example, the Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1995b) is, unsurprisingly, allied with the three level model and attempts to measure SA across the supposed three levels. SART on the other hand takes a multi-dimensional approach to SA that is congruent with Taylor's (1990) view that SA comprises the level of demand imposed on attentional resources by a situation; the supply of attentional resources in response to these situational demands; and the subsequent understanding of the situation.

### **Situation Awareness Measures**

In a review of SA measurement techniques, Endsley (1995b) describes a range of different approaches that have been used in the past, including physiological measurement techniques (e.g. eye tracking devices, electroencephalograms etc), performance measures (e.g. external task performance measures and imbedded task performance measures), subjective rating techniques (self and observer rating), questionnaires (post-trial and on-line) and freeze probe techniques (e.g. SAGAT). A recent review of human factors methods identified over twenty different approaches designed specifically for the measurement of SA (Stanton, Salmon, Walker, Baber & Jenkins, 2005). In a review of SA measurement approaches, Salmon, Stanton, Walker & Green (2006) identified the following categories of SA assessment methods: freeze probe recall techniques, real-time probe techniques, post trial subjective rating

techniques, observer rating techniques, process indices and performance measures. A brief summary of each type of measurement approach is given below.

### *Freeze probe techniques*

Freeze probe techniques involve the administration of SA related queries on-line during 'freezes' in a simulation of the task under analysis. Typically, a task is randomly frozen, all displays and screens are blanked and a set of SA queries regarding the current situation at the time of the freeze is administered. Participants are required to answer each query based upon their knowledge and understanding of the situation at the point of the freeze. Participant responses are compared to the state of the system at the point of the freeze and an overall SA score is calculated at the end of the trial. SAGAT (Endsley, 1995b) is the most popular freeze probe technique and was developed to assess pilot SA based on the three levels of SA proposed in the Endsley's three-level model. SAGAT uses queries of designed to assess participant SA, including level 1 SA (perception of the elements), level 2 SA (comprehension of their meaning) and level 3 SA (projection of future status) related queries.

### *Real-time probe techniques*

Real-time probe techniques involve the administration of SA related queries on-line (during task performance), but with no freeze of the task under analysis. Typically, SMEs develop queries either prior to the task or during task performance and administer them while the participant is performing the task under analysis. Answer content and response time are taken as a measure of participant SA. It is argued that the main advantage associated with 'real-time' probe techniques is the reduced level of intrusiveness compared to freeze probe approaches, since no freeze of the task is

required. The situation present assessment method (SPAM; Durso et al 1998) is a real-time probe technique developed for use in the assessment of air traffic controllers SA. SPAM uses task-related queries to probe operators for their SA (e.g. which of the two aircraft A or B, has the highest altitude?) via telephone. The query response time (for those responses that are correct) is taken as an indicator of the operators SA and the time taken to answer the telephone acts as an indicator of workload.

### *Self-rating techniques*

Self-rating techniques are used to elicit subjective assessments of participant SA. Typically administered post-trial, self-rating techniques involve participants providing a subjective rating of their perceived SA via a rating scale of some sort. The primary advantages of self-rating techniques are their ease of application (easy, quick and low cost) and their non-intrusive nature (since they are administered post-trial). However, subjective self-rating techniques are heavily criticised for a plethora of reasons, including the various problems associated with the collection of SA data post-trial (correlation of SA with performance, poor recall etc) and also issues regarding their sensitivity. The Situation Awareness Rating Technique (SART; Taylor, 1990) is the most popular of these approaches. SART uses the following ten dimensions to measure operator SA: Familiarity of the situation, focussing of attention, information quantity, information quality, instability of the situation, concentration of attention, complexity of the situation, variability of the situation, arousal, and spare mental capacity. SART is administered post-trial and involves the participant rating each dimension on a seven point rating scale (1 = Low, 7 = High) in order to gain a subjective measure of SA.

### *Observer rating techniques*

Observer-rating techniques are most commonly used when measuring SA ‘in-the-field’ due to their non-intrusive nature. Observer rating techniques typically involve SMEs observing participants during task performance and then providing an assessment or rating of each participant’s SA. The SA ratings are based upon pre-defined observable SA related behaviours exhibited by participants during task performance. The situation awareness behavioural rating scale (SABARS) is an observer rating technique that has been used to assess infantry personnel situation awareness in field training exercises (e.g. Matthews & Beal 2002). The technique involves domain experts observing participants during task performance and rating them on 28 observable SA related behaviours. A five point rating scale (1=Very poor, 5 =Very good) and an additional ‘not applicable’ category are used. The 28 behaviour rating items were designed specifically to assess platoon leader SA (Matthews et al 2000).

### *Performance measures*

Using performance measures to assess SA involves measuring relevant aspects of participant performance during the task under analysis. Depending upon the task, certain aspects of performance are recorded in order to determine an indirect measure of SA. For example, in a military infantry exercise, performance measures may be ‘kills’, ‘hits’ or mission success or failure. When assessing driver SA, Gugerty (1997) measured hazard detection, blocking car detection, and crash avoidance during a simulated driving task.



### *Process indices*

Process indices involve recording the processes that the participants use in order to develop SA during the task under analysis. One example of using process indices to assess SA is the measurement of participant eye movements during task performance (e.g. Smolensky, 1993). An eye-tracking device can be used to measure participant fixations during the task under analysis, which can then be used to determine how the participant's attention was allocated during the task under analysis. Concurrent verbal protocol analysis (VPA) involves creating a written transcript of operator behaviour as they perform the task under analysis. The transcript is based upon the operator 'thinking aloud' as he conducts the task under analysis. VPA is used as a means of gaining an insight into the cognitive aspects of complex behaviours and is often used to indicate operator SA during task performance.

### *Team SA measures*

Interestingly, only relatively little attention has been given to the development of specific team SA measures. Team SA measures tend to focus on the levels of overall team SA and/or the degree of shared awareness between members of a team. Team SA measures can be categorised into team probe-recall techniques, observer rating team SA techniques and team task performance-based SA assessment techniques. Team probe-recall techniques (e.g. Bolstad, Cuevas, Gonzalez, & Schneider, 2005) involve the use of a SAGAT style approach in a team setting. These approaches suffer from the same criticisms that are aimed at SAGAT style approaches and are also difficult to use during real world collaborative tasks (which are difficult to freeze and are often distributed over a wide geographical area). Observer rating team SA techniques involve SME observers observing team performance and rating the level

of SA that each individual team member has and also the level of team and shared awareness. Like their individual SA assessment counterparts, these approaches suffer from doubts over their validity i.e. the extent to which observers can rate other people's internal levels of SA. The majority of team SA assessment techniques come under the umbrella of team task performance-based SA assessment techniques. Typically, responses to changes in the task and environment are used to assess how aware a team and its components are.

Of the many different forms of SA measurement approach available, the majority are belied by flaws which impact their utility when used to assess team SA. Freeze probe techniques are the most commonly used approach to assess SA and have the most validation evidence associated with them (e.g. Jones & Kaber, 2005). However, the use of freeze-probe techniques to assess SA during real world collaborative tasks is often impractical since the tasks cannot be frozen and agents are distributed over geographical distances, making it difficult to administer SA probes. Self-reporting techniques, such as the Situation Awareness Rating Technique (SART; Taylor, 1990) suffer from a host of problems, including a correlation of SA ratings with performance (Endsley, 1995b), participants failing to recall aspects of the task when they had poor SA and also the absence of specific team SA self report measures. Observer rating techniques are also beset by a number of flaws including doubts about the extent to which observers can accurately rate the internal construct of SA (Endsley, 1995b) and the difficulties faced by experimenters in assembling an observer team of sufficient size and experience. Other approaches such as performance measures and process indices (e.g. eye tracker) suffer from a similar fate i.e. the extent to which performance and process can be linked to good or poor SA.

In conclusion to their SA methods review, Salmon et al (2006) reported that, of the measures available in the literature, none were suitable for using (in isolation) in the assessment of individual and team SA in command and control environments.

Further, Salmon et al (2006) reported that the SAGAT (Endsley, 1995b) and the SART (Taylor, 1990) approaches are by far the most commonly applied during individual and team SA assessments. Many researchers, however, argue that further investigation into the measurement of SA in complex systems is required, particularly with regard to team SA (e.g. Artman, 2000; Gorman, Cooke & Winner, 2006, Patrick, James, Ahmed & Halliday, 2006; Salmon, Stanton, Walker and Green, 2006; Siemieniuch & Sinclair, 2006).

### **Validating Situation Awareness Measures**

The issue of validity is always pertinent when measuring human factors issues within complex systems. Stanton & Young (1999, 2003) point out that there is little evidence that human factors methods actually work, and the way in which to ensure that they work is to assess their reliability and validity. The validity of SA measures relates to the degree to which the measure is actually measuring SA and not some other psychological construct. Uhlarik & Comerford (2002) describes the following categories that should be considered when assessing the validity of SA measurement approaches:

1. Face validity – the degree to which the measure appears to measure SA as judged by a subject matter expert.
2. Construct validity – the degree to which the measure is underpinned by a sound theory or model of SA.

3. Predictive validity – the degree to which the measure can predict SA.
4. Concurrent validity – the degree to which the measure correlates with other measures of SA.

Endsley (1995b) reports that when considering SA measurement techniques, it is necessary to establish that the technique measures SA and not some other construct, possess sufficient sensitivity to allow it to detect changes in SA and does not impact SA in any way during the measurement procedure. Further, Endsley (2000) discusses the following attributes related to the reliability and validity of the SAGAT approach: sensitivity; criterion validity; reliability; construct validity; and intrusiveness.

Of the many SA measures available, the SAGAT approach has the most encouraging validation evidence associated with it. Jones & Kaber (2005) report that evidence from numerous studies that have been conducted in order to assess the validity of SAGAT suggests that the method is a valid metric of SA. Endsley (2000) also reports that the SAGAT technique has been shown to have a high degree of validity and reliability for measuring SA.

Only a handful of studies, however, have sought to compare the different SA measurement approaches available. Endsley, Sollenberger & Stein (2000) compared the sensitivity and validity of SAGAT, a real-time probe approach and SART when used to assess air traffic controller SA. Participant SA was measured when using a traditional ATC display and an enhanced ATC display. In conclusion, Endsley, Sollenberger & Stein (2000) report that SART and the on-line probes approach did not show a significant difference in SA between conditions, whereas the SAGAT

scores were sensitive to display changes. In addition, no statistically significant correlation was found between total SAGAT scores and performance, however, significant relationships between the individual SAGAT queries and performance were identified. Moderate correlations between the different SA measures were also identified. Level 1 SAGAT scores were correlated with the overall SART rating, the supply dimension rating and the Level 1 real time probe. However, level 2 and 3 SAGAT scores correlated negatively with the SART understanding and supply ratings. Endsley, Selcon, Hardiman & Croft (1998) compared SAGAT and SART when used to assess fighter pilot SA. In this case no correlation between the SAGAT and SART measures was found. In conclusion, Endsley et al (1998) reported that “the subjective assessment of SA derived via SART does not appear to be related to the objective measure of SA provided by SAGAT (Endsley et al, 1998). Jones & Endsley (2000) compared a real time probe measure, SAGAT, SART and the NASA TLX workload measure during air sovereignty team’s performance of a low and a high workload task in the North American Aerospace Defence Regional Sector Air Operations Centre simulator. In conclusion, Jones & Endsley reported that all three SA measures demonstrated sensitivity to the differences in the two scenarios undertaken by the teams and that there was a significant correlation only between the real time probe measure and the SAGAT measure. A relationship between the NASA TLX and SART measures was also identified. Indeed, these two measures have previously been found to be highly correlated (e.g. Selcon, Taylor and Koritsas, 1991; cited in Endsley & Jones, 2000).

It seems then, based on the literature, that the concurrent validity of SA measures is limited. In the studies reported above only Endsley, Sollenberger & Stein report a

statistically significant correlation between level 1 SAGAT query and SART and level 1 real time query scores. This leads us to believe that the different forms of SA measures discussed may be actually assessing different things. To investigate this further we wished to compare SA measures in order to identify how their views on SA differed.

### *SA Measures Comparison*

The study compared a SAGAT style probe recall measure, the SART questionnaire and a CDM (Klein & Armstrong, 2005) interview to assess participant SA. A brief overview of each approach is presented below.

### *SAGAT*

A variation of the SAGAT approach was used to assess SA during the planning task. An initial SA requirements analysis was undertaken in order to identify the different SA elements involved. This involved developing a Hierarchical Task Analysis (HTA; Stanton, 2006) for the planning task and then identifying SA requirements from the HTA. Six sets of multi-question SA probes were subsequently developed. Examples of the SA elements included in the probes are presented in Table 1.

Table 1. Example SA probes

<b>Level 1 SA Elements</b>	<b>Level 2 SA Probes</b>	<b>Level 3 SA Probes</b>
Location and number of: Schools Helicopters Buildings Fires Wide Roads Sensitive Areas Building Heights (above or below 5 storeys) Armoured vehicles Foot platoons Enemy	Areas on the battlefield where it would be the most difficult to evacuate civilians  Most vulnerable locations	Armoured vehicles Foot platoons  Potential routes to Target Areas (based on resources available)

## SART

The SART approach is a subjective rating technique. In this case, participants were asked to complete the SART questionnaire upon completion of the planning task. The SART questionnaire is presented in Figure 1.

SITUATION AWARENESS RATING TECHNIQUE (SART; Taylor, 1990)

**Instability of Situation**  
How changeable is the situation? Is the situation highly unstable and likely to change suddenly (High) or is it very stable and straightforward (Low)?

1 2 3 4 5 6 7

**Complexity of Situation**  
How complicated is the situation? Is it complex with many interrelated components (High) or is it simple and straightforward (Low)?

1 2 3 4 5 6 7

**Variability of Situation**  
How many variables are changing within the situation? Are there a large number of factors varying (High) or are there very few variables changing (Low)?

1 2 3 4 5 6 7

**Arousal**  
How aroused are you in the situation? Are you alert and ready for activity (High) or do you have a low degree of alertness (Low)?

1 2 3 4 5 6 7

**Concentration of Attention**  
How much are you concentrating on the situation? Are you concentrating on many aspects of the situation (High) or focussed on only one (Low)?

1 2 3 4 5 6 7

**Division of Attention**  
How much is your attention divided in the situation? Are you concentrating on many aspects of the situation (High) or focussed on only one (Low)?

1 2 3 4 5 6 7

**Spare Mental Capacity**  
How much mental capacity do you have to spare in the situation? Do you have sufficient to attend to many variables (High) or nothing to spare at all (Low)?

1 2 3 4 5 6 7

**Information Quantity**  
How much information have you gained about the situation? Have you received and understood a great deal of knowledge (High) or very little (Low)?

1 2 3 4 5 6 7

**Familiarity with Situation**  
How familiar are you with the situation? Do you have a great deal of relevant experience (High) or is it a new situation (Low)?

1 2 3 4 5 6 7

Figure 1. SART Questionnaire

## Critical Decision Method

The CDM approach has been used by the authors to derive information related to SA during collaborative task performance (e.g. Stanton et al 2005; Stanton et al, 2006). The CDM procedure in this case involved interviewing participants post trial using the probes presented in Table 2.

Table 2. CDM Probes (Source: O'Hare et al, 2000)

Goal Specification	What were your specific goals at the various decision points?
Cue Identification	What features were you looking for when you formulated your decision? How did you that you needed to make the decision?, How did you know when to make the decision?
Expectancy	Were you expecting to make this sort of decision during the course of the event? Describe how this affected your decision making process.
Conceptual	Are there any situations in which your decision would have turned out differently?
Influence of uncertainty	At any stage, were you uncertain about either the reliability of the relevance of the information that you had available?
Information integration	What was the most important piece of information that you used to formulate the decision?
Situation Awareness	What information did you have available to you at the time of the decision?
Situation Assessment	Did you use all of the information available to you when formulating the decision? Was there any additional information that you might have used to assist in the formulation of the decision?
Options	Were there any other alternatives available to you other than the decision you made?
Decision blocking - stress	Was their any stage during the decision making process in which you found it difficult to process and integrate the information available?
Basis of choice	Do you think that you could develop a rule, based on your experience, which could assist another person to make the same decision successfully?
Analogy/ generalisation	Were you at any time, reminded of previous experiences in which a similar/different decision was made?

The CDM data is then used to construct *propositional networks*, which are a form of mind map depicting the information comprising SA for a particular task or system. Using a simple content analysis, keywords are extracted from the CDM interview responses and linked together (based on their causal links) to form a network comprising the information elements that were used by the different agents during task performance. Essentially propositional networks display the information that comprises a system's DSA. These *information elements* represent what the agents 'needed to know' in order to achieve their goals during task performance and are linked based on causal links that emerged during the task under analysis (e.g. enemy 'has' location, commander 'knows' plan etc).



Before undertaking the comparison of methods study we wished to further explore how each approach viewed and measured SA (i.e. what each approach considered participant SA to actually consist of). For this purpose, a content analysis was undertaken on each of the different methods SA queries or probes. The results of the content analysis are presented in Figure 2.

SAGAT		SART		CDM	
Locations	Platoons	Situation	Activity	Goals	Info Usage
Elements	Resources	Instability	Concentration	Decisions	Info Required
Map	Routes	Changeability	Aspects	Decision Points	Decision Formulation
Schools	Assets	Stability	Focussed	Features	Alternatives
Helicopter	Targets	Straightforward	Mental Capacity	Cues	Difficulty
Buildings	Time	Change	Information	Events	Process
Storeys	Risk	Complexity	Information Quantity	Decision Making	Integration
Fire	Vulnerable	Components	Information Quality	Situations	Rules
Civilians	Items	Variability	Knowledge	Uncertainty	Experience
Situation Overlay	Number	Variables	Communication	Reliability	Assistance
Understanding	Decision	Arousal	Familiarity	Relevance	Success
Task	Arm Vehicle	Alertness	Experience	Information	Time
Evacuate				Importance	Similar Decisions
				Availability	Different Decisions

Figure 2. Content Analysis of the three SA Measures

From this analysis it is clear that there are significant differences between the different measures perspectives on SA. The SAGAT approach queries participants for the knowledge of task-specific elements, such as buildings, platoons, routes and resources and their status, such as locations, vulnerability, and number. SART, on the other hand, uses more generic, global aspects of the scenario to derive a measure of

participant SA, such as situation instability, complexity, changeability and participant alertness, concentration, knowledge and experience. SART therefore makes no reference to the specific elements related to the task, rather it focuses on generic, overall task characteristics. The CDM approach is more concerned with goals, decision making and information related to key events during task performance. This approach therefore elicits information related to critical events during the task. Whereas SAGAT gathers knowledge of specific elements and SART focuses on generic aspects of the overall task, CDM lies somewhere between the two, since it attempts to garner detailed information about different events that occurred during task performance.

The content analysis findings suggest that both the analysis requirements and the characteristics of the task under analysis should drive the SA measure that is used. If the task and outcome is well known and stable (e.g. simulated tasks) and an analysis of how aware an individual is, then it appears that approaches such as the SAGAT approach are the most useful, since probes can be developed a priori based on the analysis requirements. In short, if the experimenter knows what SA should comprise beforehand, then SAGAT type approaches appear to be the most suitable approach to use. It is apparent then that the SAGAT approach can be classified as deterministic and linear, in that one defines what SA is and then measures it. However, if SA content is not pre-defined and the task is dynamic, collaborative, and changeable and the outcome is not known (e.g. real world tasks) then a different, non-deterministic and non-linear, approach may be more suited.

The aims of this study were therefore to compare three different SA measurement approaches when used to assess SA during a collaborative military planning task. In particular, we wished to identify which of the three measures was the most accurate at measuring SA and also we wished to assess how each approach viewed and represented the construct.

## **Methodology**

### *Design*

A mixed experimental design was used. The between subject variables were the situational overlay accuracy and participant SA (as measured by the three different approaches) and the within subject variable was the three SA scores for each participant (provided by the three approaches used). The dependent variables were time taken to complete the task, participant SA, mental workload, and situational overlay accuracy.

### *Participants*

A total of twenty participants (13 female and 7 male) were used during this study. The mean age of the participants was 30 years old (SD = 10.17). Participants were recruited via poster and email advertisement and comprised undergraduate students currently attending Brunel University.

### *Materials*

#### *Brunel Silver Command Wall*

The study was undertaken within Brunel University's command wall command and control experimental test bed environment (Green, Stanton, Walker & Salmon, 2005). The Brunel command wall system comprises a command wall interface (located at the command centre) and wearable units (worn by agents in the field). The command wall incorporates custom made wearable computers and a Wi-Fi enabled urban battlespace; a paradigm of NEC. The centre screen provides a 2D view of the battlespace, whilst the right hand side screen provides a Google Earth based 3D view of the battlespace. Both battlespace views can be manipulated using a mouse controller (i.e. zoom in and out, rotate etc). The 2D display screen has an overlay drawing function which allows the commander to draw situational overlays and colour code different areas/objects within the battlespace (e.g. red = enemy). The commander interacts with the command wall interface through standard mouse controller and keyboard devices. The command wall is linked directly to agents in the field via a wireless network and wearable technology units. Agents in the field are dynamically represented on the 2D and 3D map display by their spatial location in real time. The command wall interface is presented in Figure 3. A representation of the Brunel campus is presented in Figure 4.



Figure 3. The Brunel Silver Command Wall System

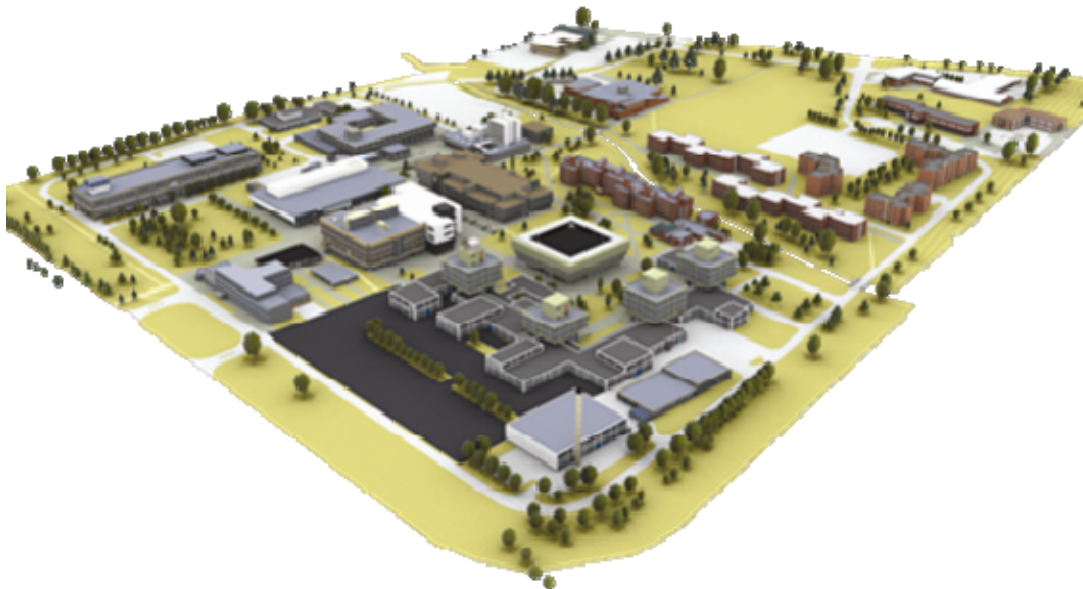


Figure 4. The Battle Space Area

The Brunel University campus provides a realistic and, therefore, ecologically valid urban battlespace landscape. The campus covers an approximately rectangular area of

50 hectares (WiFi coverage is 20 hectares), with an elevation of 7.5 metres and no significant gradient. The campus is laid out with 20 definable structures (mainly concrete) ranging in height from approximately 3 metres to 20 metres (1 story to 8 stories respectively). The land adjacent to and between the structures is covered with hard paving and grass. The total battlespace is bounded by a perimeter road on all boundary faces, beyond which is chain link fencing on the South and North boundaries, a public road on the West boundary and a small river on the East boundary.

### *The Combat Estimate*

The experiment was based around the Combat Estimate (CE) planning Technique. The CE process is a military planning process that is currently used by the UK armed forces. The study focussed on the use of question one of the CE process to develop SA of the battlespace area and the enemy and threat. Question one comprises two initial phases, the Battlefield Area Evaluation (BAE) and the threat evaluation followed by a threat integration phase. Question one involves defining the battlespace, describing the battlespace effects, evaluating the enemy and describing the enemies course of action. The BAE phase involves an assessment of the battlespace on both friendly and enemy operations. The aim of the BAE is analyse the terrain so that Mobility Corridors (MCs), Avenues of Approach (AAs) and Manoeuvre Areas (MAs), key terrain, choke points and restricted and severely restricted areas can be identified. The threat evaluation phase is used to identify the enemy's doctrinal norms. Once the BAE and threat evaluation phases are complete, threat integration is used to identify how the enemy are likely to operate during the battle.

### *Study Scenario*

The scenario involved the Commander (participant) undertaking question one of the CE process for an experimental Brunel campus Military Operations in Urban Terrain (MOUT) warfare scenario. Based on intelligence from the battlefield (provided by the field agents), participants were required to perform a BAE comprised of terrain and threat analysis tasks (i.e. "what are the enemy doing and why?"). In this case the field agents were not real but rather were simulated in order to control variations in their ability to resolve such information. The information sent to the Commander by the field agents was pre-programmed to enable information to automatically appear on the commanders map display at pre-determined intervals. The Commander's task was to use the incoming information from the field agents in order to construct a situation overlay that represented the BAE. In particular, the Commander had to highlight on the map display the following types of key terrain:

1. Buildings 3 stories or higher;
2. Routes and roads;
3. Sensitive areas; and
4. Hostile areas.

In addition, the commander was required to highlight various types of events on the map, including the locations of a sniper, barricade, road traffic accident, bomb and fire.

The following materials were used during the study: The Brunel command wall system, including the experimental environment, three screen wall display containing 2D and 3D representations of the experimental environment, and a standard keyboard and mouse. A laptop containing a PowerPoint presentation was used to administer the

SAGAT probes. SART, NASA TLX and Critical Decision Method questionnaires were also administered post trial. Two stop watches were used, the first one to time the duration of the experiment and the second one to time the duration of responses to the SAGAT probes and other questionnaires. Pen and paper were provided to participants to make notes and also to complete the consent form. A demographic questionnaire was administered to participants, along with experimental instructions.

### *Procedure*

After a short briefing, participants were given an instructions booklet and were taken through the experimental procedure in order to clarify what was required of them during the experiment. In order to familiarise them with the experimental system, participants then undertook a short trial which involved constructing a situational overlay using the command wall system and also answering a series of SAGAT probes. Next participants were asked to undertake the experiment for real. This involved them using the command wall system to construct a situational overlay based on the incoming intelligence provided by simulated field agents. SAGAT probes were administered at random points during the experiment. This involved the task being frozen, the silver command walls screens being blanked and the administration (via laptop computer) of SAGAT probes. The time taken to complete the entire experiment and also each SAGAT probe was recorded by the experimenter. Upon completion of the experimental trial, participants were asked to complete a NASA TLX workload questionnaire, a SART SA questionnaire and a CDM interview.

### **Results**



### *Situational Overlay Accuracy*

The accuracy of participants situational overlay construction was calculated by comparing their completed situational overlay against a pre-defined 'expert' overlay for the same scenario, which reflected exactly the battle field area situation that was presented to the commander. In total there were 25 markers contained in the expert situational overlay. The percentage of correct markers added to the situational overlays by each participant is presented in Figure 5. The mean percentage score was 85.4% (SD = 0.09).

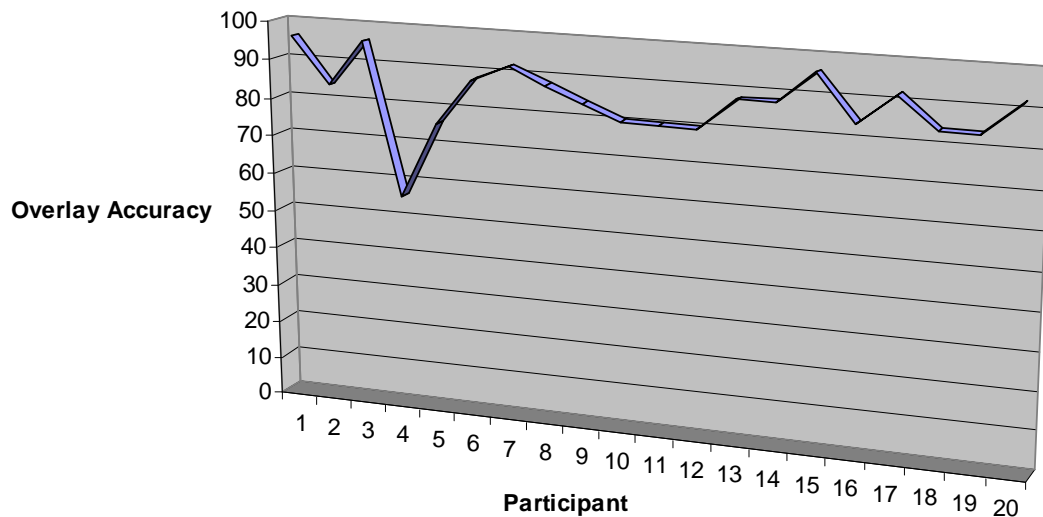


Figure 5. Participants correct response percentage for the situation overlay task.

### *SAGAT*

Participant responses to the SAGAT probes were scored either as 1 (correct) or 0 (incorrect). Participant's total SAGAT scores were calculated by summing all correct responses which gave them a possible total score of 24. Participant total SAGAT scores are presented in Figure 6. The mean total SAGAT score was 11.35 (SD =

3.77). The highest total SAGAT score was achieved by participant No.13, who scored 19. The lowest SAGAT score was achieved by participant No.18, who scored 5.

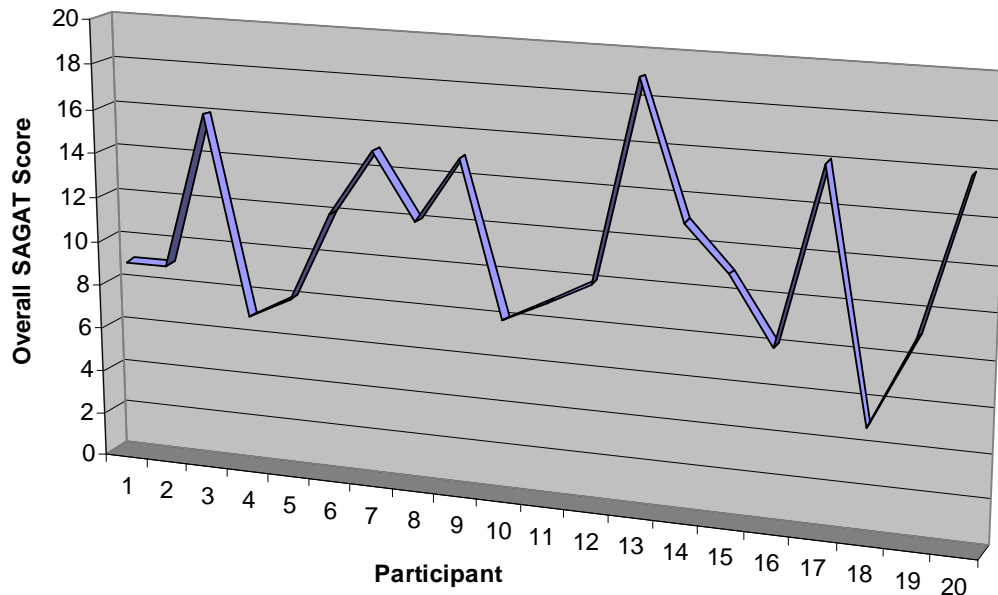


Figure 6. Participant total SAGAT scores

The SAGAT scores were also decomposed based on their corresponding SA level. The total SAGAT scores per SA level for each participant are presented in Figure 7. The mean overall participant SAGAT score for level 1 SA probes was 7 (SD = 1.87). The highest overall participant SAGAT score for level 1 SA probes was achieved by participant No.17, who scored a total of 11. The mean overall participant SAGAT score for level 2 SA probes was 3.95 (SD = 2.09). The highest overall participant SAGAT score for level 2 SA probes was achieved by participants No.3 and 13, who scored a total of 8. The mean overall participant SAGAT score for level 3 SA probes was 0.40 (SD = 0.49). The highest overall participant SAGAT score for level 3 SA probes was 1.

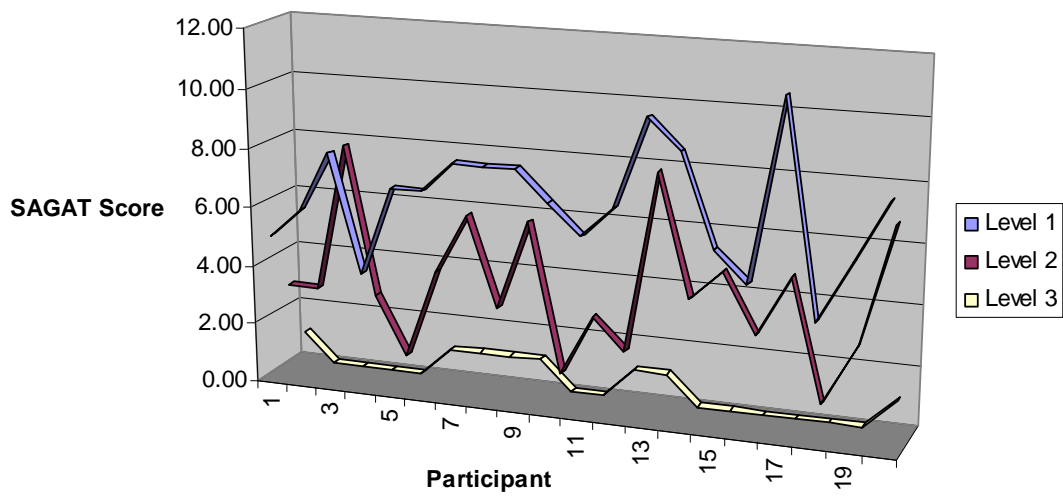


Figure 7. Total SAGAT scores per SA level

*SART*

An overall SART score was derived using the following formula:  $SA = U - (D - S)$ ,

where:

U = summed understanding

D = summed demand

S = summed supply

The overall SART scores for each participant are presented in Figure 8. The mean overall SART score was 19.75 (SD = 5.7).

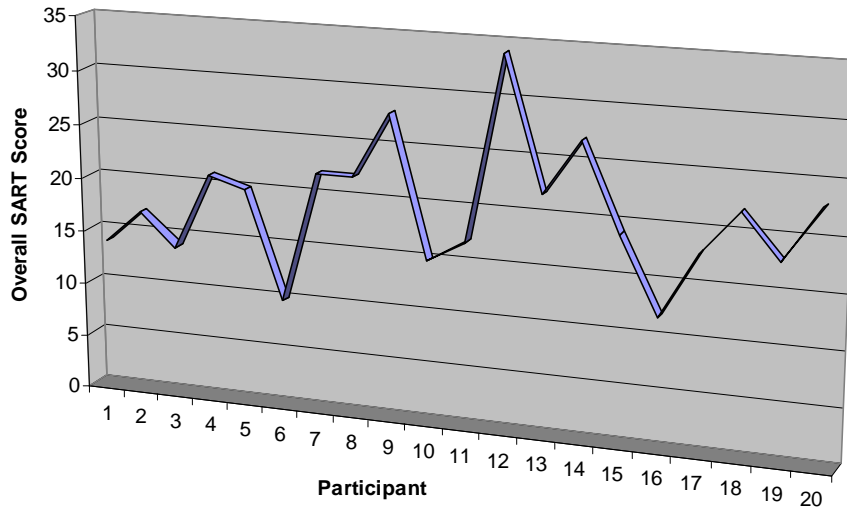


Figure 8. Overall SART scores

Participant scores for each SART dimension (*Supply*, *Demand* and *Understanding*) are presented in Figure 9. The mean score for *Demand* was 13.9 (SD = 3.95). The mean score for *Supply* was 20.15 (SD = 4.84) and the mean score for *Understanding* was 13.5 (SD = 3.03).

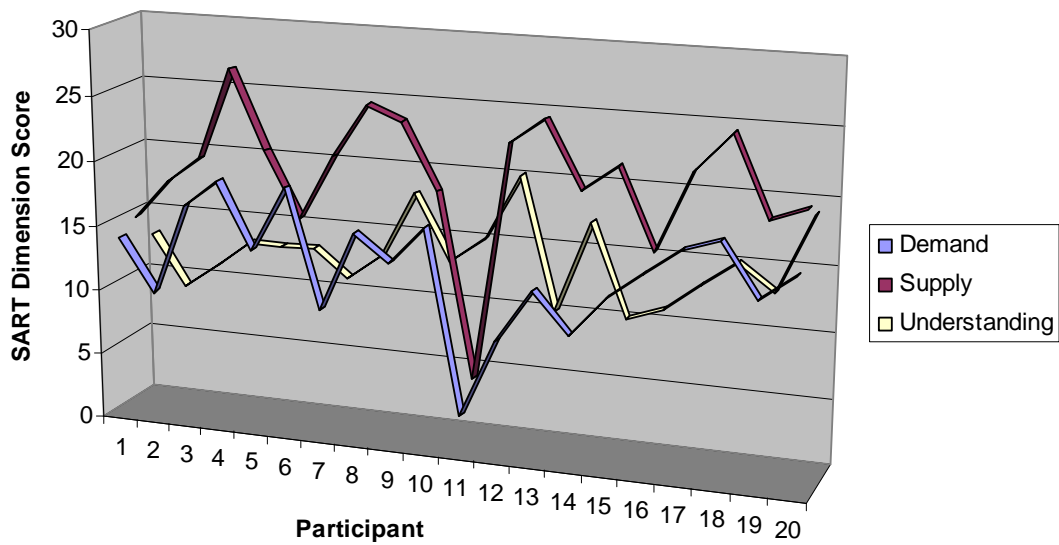


Figure 9. Participant SART dimension scores

### *Critical Decision Method*

Participant responses to the CDM questionnaire were coded and used to construct propositional networks for each participant. Participant propositional networks were then compared with an ‘expert’ propositional network for the scenario in order to assess their SA during the experiment. The signal detection paradigm was used to generate an SA score for each participant. The signal detection paradigm sorts the data into the following mutually exclusive categories:

- 1) Hit – A knowledge element reported by the participants in the post-trial questionnaire that was present in the expert propositional network.
- 2) Miss – The failure to report a knowledge element present in the expert propositional network in the post trial questionnaire.
- 3) False Alarm – A knowledge element reported by the participant that was not present within the expert propositional network.

4) Correct rejections – Correctly rejected error that was present neither in the expert propositional network nor in the total pool of false alarms made by the other participants.

These four categories were entered into the signal detection grid for each subject. The signal detection paradigm was then used to calculate a sensitivity index (SI) score for each participant. Participant SI scores were then converted into  $d'$  scores, which are presented in Figure 10. The formula used to calculate SI is given in Formula 1 (from Stanton and Stevenage, 2000).

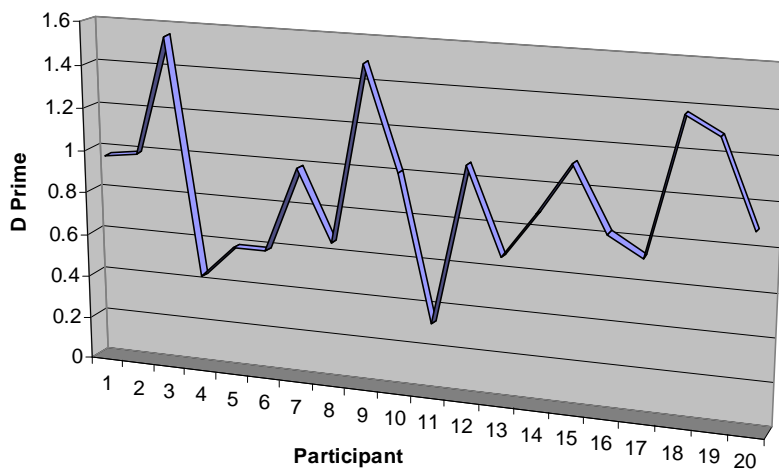


Figure 10. Participant  $d'$  Scores

$$Si = \left( \frac{\left( \frac{\text{Hit}}{\text{Hit} + \text{Miss}} \right) + 1 - \left( \frac{\text{False Alarm}}{\text{FA} + \text{Correct Rejection}} \right)}{2} \right)$$

Formula 1. Sensitivity Index formula

## Correlations

Spearman's Rho statistical tests were undertaken to identify significant correlations between the following variables:

1. Performance and SA (as measured by the three different approaches);
2. Performance and Workload (as measured by the NASA TLX);
3. Between the different SA measures (e.g. SAGAT versus SART, SAGAT versus Prop Nets and SART versus the Prop Nets) and between the different dimensions measured (e.g. SAGAT levels 1, 2 & 3 and SART understanding, SART demand and SART supply).

The statistical analysis findings are presented in Tables 3, 4, 5 and 6.

## Performance and SA, Workload and Time

The correlation between participant overall SAGAT SA scores and performance was significant (.662, <0.01). There was also a significant correlation (.691, <0.01) between level 2 query SAGAT scores and performance. There were no significant correlations between the other SAGAT scores (levels 1 and 3), SART (Total, Understanding, Supply and Demand) and CDM d' scores and performance. The performance and workload scores comparison produced a correlation co-efficient of

.330 which was not significant. There was a significant negative correlation between time and performance (-.474, <0.05) which suggests that the participants who took less time performed better in the overlay accuracy test. The correlation co-efficient and their corresponding significance levels are presented in Table 3.

Table 3. Correlation between Performance and SA, Workload and Time.

	SAGAT	SAGAT Level 1	SAGAT Level 2	SAGAT Level 3	SART	SART U	SART D	SART S	CDM	Workload	Time
Performance	.662 <0.01	.385 <i>Not Sig</i>	.691 <0.01	.297 <i>Not Sig</i>	-.120 <i>Not Sig</i>	-.120 <i>Not Sig</i>	-.120 <i>Not Sig</i>	-.120 <i>Not Sig</i>	.273 <i>Not Sig</i>	.330 <i>Not Sig</i>	-.474 <0.05

### SA Measures

The analysis of the correlations between the different SA measures used is presented in Tables 5 (SAGAT and SART), 6 (SAGAT and CDM) and 7 (SART and CDM).

There were no significant correlations between the participant SA scores provided by the three different SA measures as shown in tables 4, 5 and 6.



Table 4. Correlations between SAGAT and SART measures

	SART	SART U	SART D	SART S
SAGAT	.274 <i>Not Sig</i>	-.044 <i>Not Sig</i>	-.152 <i>Not Sig</i>	.143 <i>Not Sig</i>
SAGAT Level 1	.375 <i>Not Sig</i>	-.230 <i>Not Sig</i>	-.124 <i>Not Sig</i>	.230 <i>Not Sig</i>
SAGAT Level 2	.123 <i>Not Sig</i>	-.230 <i>Not Sig</i>	-.052 <i>Not Sig</i>	.101 <i>Not Sig</i>
SAGAT Level 3	.356 <i>Not Sig</i>	.220 <i>Not Sig</i>	-.107 <i>Not Sig</i>	.160 <i>Not Sig</i>

Table 5. Correlations between SAGAT and CDM

	CDM
SAGAT	.095 <i>Not Sig</i>
SAGAT Level 1	-.023 <i>Not Sig</i>
SAGAT Level 2	.115 <i>Not Sig</i>
SAGAT Level 3	-.035 <i>Not Sig</i>

Table 6. Correlations between SART and CDM

	CDM
SART	.154 <i>Not Sig</i>
SART U	.009 <i>Not Sig</i>
SART D	-.045 <i>Not Sig</i>
SART S	.089 <i>Not Sig</i>

## Discussion

### *Performance*

Participant performance on the situational overlay construction task was encouraging, with the majority of participants (18 out of 20) achieving an accuracy score of 80% or above. The mean score was 85.4% (SD = 0.09). These findings suggest that the majority of participants were able to successfully complete a military planning task using the Brunel command wall interface.

### *Comparison of SA measures*

Of the three methods used, only the overall and level 2 SAGAT scores produced a statistically significant correlation with performance. This indicates that the higher the participant's overall and level 2 SAGAT scores were, the more accurate they were in the situation overlay construction task. It is concluded from this that, of the methods tested, the SAGAT approach was the most accurate at measuring participant SA during the study. Since the BAE task analysed involved constructing a situation overlay which represented the current situation on the battlefield, it was taken that

levels of SA should correspond with levels of performance. This data therefore offers encouraging support for the SAGAT tool when used to measure SA during a BAE task undertaken in a simulated environment. Despite this conclusion, appropriate caution is recommended when interpreting these results since (as many researchers have pointed out) the link between SA and performance is tenuous. Good SA does not necessarily guarantee successful task performance, whilst poor SA does not necessarily guarantee poor task performance. Endsley (2000), for example, points out that, although SA is key to decision making, many other factors are involved in turning good SA into successful performance and it is possible to make wrong decisions with perfect SA and good decisions with poor SA.

This finding suggests then that, when the task environment is relatively stable and the SA-related elements and associated states and properties can be accurately identified prior to task performance (as was the case with this study) it is appropriate to use a SAGAT type approach to measure participant SA. The ability to accurately define the SA elements related to a task a priori therefore enhances the utility of the SAGAT approach as a measure of SA.

There was no significant correlation between the participant SA scores provided by the three SA measures used. It is concluded from this that the different methods view the construct of SA differently and were essentially measuring different elements of the participant's awareness during the study. SAGAT, a probe recall approach, essentially measures the extent to which a participant is aware of a pre-defined element in the environment, their understanding of the properties of these elements in relation to the task they are performing, and also what the potential future states of these elements might be. SART provides a measure of how aware participant's

perceived themselves to be during task performance (based on ratings of understanding, supply and demand) and makes no reference to the different elements within the environment. Finally, the CDM method presents a subjective description of goal-related decision making during task performance, from which information elements are extracted. This presents a description of the participant's subjective view of the systems SA in terms of information elements. Each method therefore takes a different view on what SA is and what it comprises and, as the lack of a correlation between the measures indicates, these findings suggest that the different methods are measuring different things when assessing participant SA.

The findings derived from this study can be compared to previous research that has compared SA measures. When using SAGAT and SART to assess fighter pilot SA, Endsley, Selcon, Hardiman & Croft (1998) found no correlation between SAGAT and SART scores (which was demonstrated in this study), whilst Endsley, Sollenberger & Stein (2000) reported that there was a significant correlation between Level 1 SAGAT scores and overall SART ratings. Endsley, Sollenberger & Stein also reported that there was no correlation between total SAGAT scores and performance, whereas in this study a significant correlation between overall and level 2 SAGAT scores and performance was identified. Further, Endsley, Sollenberger & Stein (2000) reported that there was a significant correlation between Level 1 SAGAT scores and overall SART ratings, the SART supply dimension ratings and the Level 1 real time probes. Jones & Endsley (2000) reported a significant correlation between SART and the NASA TLX workload ratings. This was not identified during this study.

The poor performance of the CDM-based approach in this study is surprising. This approach has been used by the authors to describe Distributed Situation Awareness (DSA) in a number of studies on command and control SA in different complex sociotechnical systems (e.g. Stanton et al, 2006; Walker, Gibson, Stanton, Baber, Salmon & Green, 2006), the outputs of which have been validated by subject matter experts as providing accurate descriptions of SA. It may be that the participant's lack of exposure to such approaches impacted their responses during CDM data collection. Further, the findings indicate that a more suitable set of CDM interview probes may be required.

In terms of the application of the different methods, it was concluded from this that the SART approach was the easiest to apply and required the least in terms of additional materials and resources invested. The SAGAT approach in this case required considerable preparative activities (e.g. SA requirements analysis, development of probes etc) whilst the CDM required a lengthy data analysis procedure (e.g. content analysis, construction of prop nets etc). The SAGAT and CDM also required additional materials (e.g. simulator, WESTT software), whereas SART did not.

#### *Implications for SA measurement*

The findings from this study (and from previous research) raise doubts over the validity of SART and CDM as measures of SA. Since SAGAT was correlated with performance, but SART and CDM was not correlated with either SAGAT or performance, questions are raised over the construct validity of SART and CDM. The additional qualitative content analysis, however, leads us to conclude that the characteristics of the task and analysis should be used to inform the selection of an

appropriate SA measure. In conclusion then, the findings from this study indicate that the three SA measures compared are effectively measuring different aspects of SA. In terms of implications for the selection of SA measures, the findings suggest that, since SAGAT is deterministic and linear in nature, it is of most use when analysts can identify SA elements a priori. On the other hand, in complex real world activities where SA cannot meaningfully be defined beforehand (e.g. military planning and execution activities) and where the outcomes are not easily predictable, alternative approaches may be required. In previous real-world SA assessments, the authors have used the CDM-based approach (e.g. Stanton et al, 2006; Walker et al, 2006) since it allows the people involved to describe what SA comprised post trial, the content of which can then be validated by subject matter experts. However, the findings of this study indicate that a more suitable set of CDM interview probes may be required. The application of SA measures based on task characteristics is presented in Figure 11.

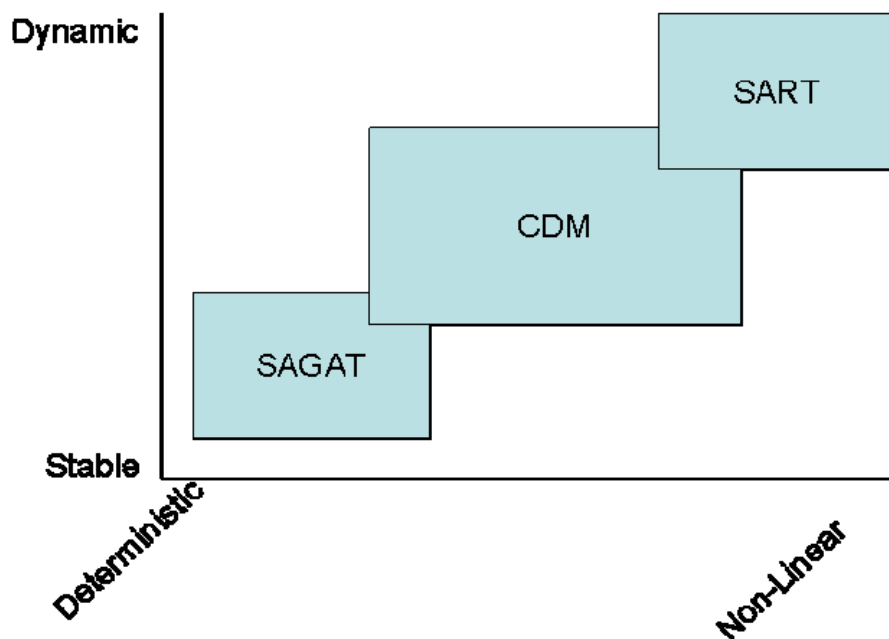


Figure 11. Relationship between task and system characteristics and SA measures

It is also worth noting that each of the methods differing perspectives on SA mean that it may be worthwhile to use a battery of SA measures when assessing the construct in complex collaborative environments. When assessing SA, there are various features that experimenters and researchers wish to focus on, including how aware an individual or team are, what their SA is comprised of, where SA-related information is coming from and how it is distributed around a system to name but a few. It is clear that no one SA measure can provide all of these assessments and so it may be that a battery of different but compatible measures be used. One such battery of measures might include the SAGAT approach (to assess how aware individuals are about elements in the environment), CDM (to identify the information that SA comprises and how it is distributed around a team and system), social network analysis (to assess the communication of SA-related information between agents) and a subjective assessment approach (to identify how aware agents felt they were during task performance).

In closing, the findings from this study suggest that SAGAT is the most suitable approach for assessing SA in tasks where the environment is stable and where SA-related elements can be pre-defined. Further, the findings indicate that the three SA measures SAGAT, SART and CDM view the construct from differing perspectives and so measure the construct in different ways. It is therefore important that experimenters select SA measures that are congruent with the way in which they view the construct and also that their selection is driven by the characteristics of the task under analysis and the nature of the research question.

## **Acknowledgement**

This research from the Human Factors Integration Defence Technology Centre was part funded by the Human Sciences Domain of the UK Ministry of Defence Scientific Research Programme.

## **References**

- Artman, H., 2000, Team situation assessment and information distribution.  
*Ergonomics*, Vol 43, pp 1076 – 1095.
- Bell, H. H. & Lyon, D. R., 2000, Using observer ratings to assess situation awareness,  
In M. R. Endsley (Ed) *Situation awareness analysis and measurement*,  
Mahwah, NJ: Laurence Erlbaum Associates.
- Bolstad, C. A., Cuevas, H. M., Gonzalez, C., & Schneider, M., 2005, Modeling shared  
situation awareness. Paper presented at the 14<sup>th</sup> Conference on Behaviour  
Representation in Modeling & Simulation (BRIMS), Los Angeles, CA.
- Durso, F.T., Hackworth, C.A., Truitt, T., Crutchfield, J., Manning, C.A., 1998.  
*Situation awareness as a predictor of performance in en route air traffic  
controllers*. *Air Traffic Quarterly*, 6, pp. 1-20.
- Endsley, M.R., 1995a, Towards a theory of situation awareness in dynamic systems,  
*Human Factors*, Vol. 37, pp. 32-64.
- Endsley, M.R., 1995b, Measurement of situation awareness in dynamic systems,  
*Human Factors*, Vol. 37, pp. 65-84.



- Endsley, M.R., 2000. Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley and D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum Associates.
- Endsley, M. R., Sollenberger, R., & Stein, E., 2000. Situation Awareness: A comparison of measures. In Proceedings of the Human Performance, Situation Awareness and Automation: User-Centred Design for the New Millenium. Savannah, GA.
- Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G., 1998, A comparative evaluation of SAGAT and SART for evaluations of situation awareness. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (pp. 82-86). Santa Monica, CA: Human Factors and Ergonomics Society.
- Fracker, M., 1991. Measures of Situation Awareness: Review and Future Directions (Rep. No.AL-TR-1991-0128). Wright Patterson Air Force Base, Ohio: Armstrong Laboratories, Crew Systems Directorate.
- Green, D., Stanton, N., Walker, G., & Salmon, P. (2005). Using wireless technology to develop a virtual reality command and control centre. *Virtual Reality*, Vol 8, pp. 147 - 155.
- Gorman, J. C., Cooke, N., & Winner, J. L., 2006. Measuring team situation awareness in decentralised command and control environments. *Ergonomics*, Vol 49, pp 1312 – 1326.
- Gugerty, L.J., 1997, Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory, *Journal of Experimental Psychology: Applied*, 3, pp. 42-66.

- Jones, D. G., & Endsley, M. R., 2000. Can real-time probes provide a valid measure of situation awareness? In Proceedings of the Human Performance, Situation Awareness and Automation: User-Centred Design for the New Millennium. Savannah, GA.
- Jones, D.G., and Kaber, D.B., 2004. In N. Stanton, Hedge, Hendrick, K. Brookhuis, E. Salas (Eds.) *Handbook of Human Factors and Ergonomics Methods*. Boca Raton, USA, CRC Press.
- Klein, G. & Armstrong, A. A., 2005. Critical decision method. In N. A. Stanton, A. Hedge, E. Salas, H. Hendrick, & K. Brookhaus, (Eds.), *Handbook of Human Factors and Ergonomics Methods*. Boca Raton, Florida, CRC Press, (pp. 35.1-35.8).
- Matthews, M.D., Beal, S. A., 2002. Assessing Situation Awareness in Field Training Exercises. U.S. Army Research Institute for the Behavioural and Social Sciences. Research Report 1795.
- Moray, N. 2004. Ou' sont les neiges d' antan? Proceedings of the HPSAII. LEA: Mahwah, NJ.
- Niesser, U., 1976, *Cognition and reality: Principles and implications of cognitive psychology*. Freeman, San Francisco.
- O'Hare, D., Wiggins, M., Williams, A. and Wong, W., 2000. Cognitive task analysis for decision centred design and training. In J. Annett and N.A. Stanton (Eds.) *Task Analysis*. London: Taylor and Francis (pp. 170-190).
- Patrick, J., James, N., Ahmed, A., and Halliday, P., 2006. Observational assessment of situation awareness, team differences and training implications. *Ergonomics*, Vol 49, pp 393 – 417.

- Riley, J. M., Endsley, M. R., Bolstad, C. A., & Cuevas, H. M., 2006. Collaborative planning and situation awareness in army command and control. *Ergonomics*, Vol 49, pp 1139 – 1153.
- Salmon, P., Stanton, N., Walker, G., & Green, D., 2006. Situation awareness measurement: A review of applicability for C4i environments. *Journal of Applied Ergonomics*, 37, 2, pp. 225-238.
- Shu, Y., & Furuta, K., 2005. An inference method of team situation awareness based on mutual awareness. *Cognition Technology & Work*, 7, pp272–287
- Siemieniuch, C. E., & Sinclair, M. A., 2006. Systems Integration. *Journal of Applied Ergonomics*, 37, pp. 91-110.
- Smith, K., & Hancock, P.A., 1995. Situation awareness is adaptive, externally directed consciousness, *Human Factors*, 37, 1, pp. 137-148.
- Smolensky, M. W., 1993. Toward the physiological measurement of situation awareness: The case for eye movement measurements. In Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, Santa Monica, Human Factors and Ergonomics Society.
- Stanton, N. A., 2006. Hierarchical task analysis: Developments, applications and extensions. *Applied Ergonomics*, 37, pp. 55-79.
- Stanton, N.A., Young, M., 1999. What price ergonomics? *Nature* 399, pp. 197-198.
- Stanton, N.A., Young, M., 2003. The application of ergonomics methods by novices. *Applied Ergonomics* 34, pp. 479-490.
- Stanton, N.A., & Stevenage, S.V., 2000. Learning to predict human error: issues of acceptability, reliability and validity. In J. Annett & N. A. Stanton (Eds.) *Task Analysis*. Taylor and Francis, London.

- Stanton, N.A., Chambers, P.R.G., Piggott, J., 2001. Situational Awareness and safety. *Safety Science*, 39, pp189-204.
- Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C. & Jenkins, D., 2005. Human factors methods: A practical guide for engineering and design. Ashgate: Aldershot.
- Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., Salmon, P. M., Hoyle, G., Walker, G. H., Young, M. S., Linsell, M., Dymott, R., & Green, D., 2006, Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. *Ergonomics*, Vol 49, pp 1288 – 1311.
- Taylor, R.M., 1990. Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. In *Situational Awareness in Aerospace Operations (AGARD-CP-478)* pp3/1 –3/17, Neuilly Sur Seine, France: NATO-AGARD.
- Uhlarik, J. and Comerford, D. A., 2002. A review of situation awareness literature relevant to pilot surveillance functions. (DOT/FAA/AM-02/3) Washington, DC: Federal Aviation Administration, U.S. Department of Transportation.
- Walker, G. H., Gibson, H., Stanton, N. A., Baber, C., Salmon, P. M., & Green, D., 2006, Event analysis of systemic teamwork (EAST): a novel integration of ergonomics methods to analyse C4i activity. *Ergonomics*, Vol 49, pp 1345 – 1369.