



Defence Research and  
Development Canada

Recherche et développement  
pour la défense Canada



# Measurement of individual and team situation awareness: A critical evaluation of the available metrics and tools and their applicability to command and control environments

*R. Breton  
DRDC Valcartier*

*S. Tremblay  
Université Laval*

*S. Banbury  
B4H Ltd. U.K.*

**Defence R&D Canada – Valcartier**

Technical Report

DRDC Valcartier TR 2003-283

May 2007

Canada



# **Measurement of individual and team situation awareness: A critical evaluation of the available metrics and tools and their applicability to command and control environments**

R. Breton  
DRDC Valcartier

S. Tremblay  
Université Laval

S. Banbury  
B4H Ltd. U.K.

**Defence R&D Canada - Valcartier**

Technical Report

DRDC Valcartier TR 2003-283

May 2007

Author/Auteur

---

Richard Breton

Approved by / Approuvé par

---

Eloi Bossé

Section Head, DSS/ Chef de Section, SAD

Approved for release by / Approuvé pour distribution par

---

Christian Carrier

Chief Scientist/ Scientifique en chef

This report follows from a psychometric evaluation of the SA measurement tools performed by Dr Sébastien Tremblay within the framework of a contract with DRDC Valcartier (W7701-1-3387) under the scientific authority of Dr Richard Breton. This work was carried out under work unit 11ba20.

© Her Majesty the Queen as represented by the Minister of National Defence, 2007

© Sa majesté la reine, représentée par le ministre de la Défense nationale, 2007

## Abstract

---

The objectives with this document are to provide a critical evaluation of the psychometric properties of the available individual and team situation awareness (SA) measurement tools and techniques and to evaluate the applicability of the SA measurement tools found in the literature to C2 environments. This report is based on an exhaustive literature survey of SA measurement and a search for tools and techniques that have been used to measure individual and team SA. The emphasis is on the psychometric properties of the tools, such as reliability, validity and sensitivity. The survey reveals a wide variety of tools, techniques and metrics but very little information about their psychometric properties. In many cases validity and reliability data are rather inconclusive and incomplete with rather small samples. Nevertheless some well-established measures of SA such as SAGAT and SART have proved to be useful in evaluating new interfaces or system designs and training programs. One key problem with SA measurement is the wide variety of measures for a large number of situations. There are nearly as many different measurement techniques as there are different situations in which an operator is at work and each situation has its own sets of procedures, task requirements and demands. According to the distribution of the SA measurement tools among the C2 environments, the SA concept can be measured in every situation. In some occasions, a slight modification of an existing measure can be brought to adapt that tool to another situation.

## Résumé

---

Les objectifs poursuivis par ce document sont de fournir une critique évaluative des propriétés psychométriques des outils et techniques mesurant la prise de conscience de la situation (SA) individuelle ou en équipe et de la possibilité d'application de ces outils répertoriés aux différents environnements de C2. Ce document est basé sur un relevé littéraire exhaustif de la mesure de SA et sur une recherche d'outils et méthodes utilisés pour mesurer ce concept. L'accent est mis sur les propriétés psychométriques d'outils telles que la fidélité, validité et sensibilité. Le relevé littéraire révèle une grande variété d'outils, techniques et métriques, mais peu d'information concernant leurs propriétés psychométriques. Dans plusieurs cas, les données concernant la validité et la fidélité sont non concluantes, incomplètes et basées sur des échantillons restreints. Néanmoins, on peut retrouver des mesures de SA bien établies telles que SAGAT et SART qui ont démontré leur utilité en relation avec l'évaluation ou le design de nouvelles interfaces et le développement de nouveaux programmes d'entraînement. L'un des problèmes liés à la mesure est qu'on peut presque associer une technique différente de mesure à chaque situation potentielle dans laquelle un opérateur est impliqué. Chacune de ces situations possède son propre ensemble de procédures, de requis et de demandes liés à la tâche. Selon la répartition des outils de mesure de SA parmi les différents environnements de C2, le concept de SA peut être mesurer dans toutes ces situations. Dans certaines circonstances, une légère modification à une mesure peut être apportée afin de l'adapter à une autre situation.

This page intentionally left blank.

## Executive summary

---

A first objective of the current document is to provide a critical evaluation of the psychometric properties of the available individual and team situation awareness (SA) measurement tools and techniques. This information is essential to selecting an appropriate metric for measuring SA in a given situation and being aware of the limits of that measure. This report is based on an exhaustive literature survey of individual and team SA measurement and a search for tools and techniques that have been used to measure SA. The emphasis is on the psychometric properties of the tools, such as reliability, validity and sensitivity. The survey reveals a wide variety of tools, techniques and metrics but very little information about their psychometric properties. In many cases validity and reliability data are rather inconclusive and incomplete with rather small samples. However, there are practical reasons for the lack of psychometric data: the small number of experts for a given situation (air traffic controllers, jet pilots or C2 operators) and the complexity and high cost of testing in simulated or real work environments (e.g., cockpits, C2 centers).

A second objective is to evaluate the applicability of the SA measurement tools found in the literature to C2 environments. One key problem with SA measurement is the wide variety of measures for a large number of situations. There are nearly as many different measurement techniques as there are different situations in which an operator is at work and each situation has its own sets of procedures, task requirements and demands. According to the distribution of the SA measurement tools among the C2 environments, the SA concept can be measured in every situation. In some occasions, a slight modification of an existing measure can be brought to adapt that tool to another situation.

Another problem to consider is the lack of consensus among researchers and human factor specialists on the definition and model of SA. Thus, there is no well-established and universal construct upon which a test can be developed and validated. Nevertheless some well-established measures of SA such as SAGAT and SART have proved to be useful in relation to the evaluation of new interface or design and training programs.

Breton, R, Tremblay, S., and Banbury, S. 2006. Measurement of individual and team situation awareness: A critical evaluation of the available metrics and tools and their applicability to Command and Control environments. DRDC Valcartier TR 2003-283. Defence R&D Canada - Valcartier.

## Sommaire

---

Un premier objectif poursuivi par ce document est de fournir une critique évaluative des propriétés psychométriques des outils et techniques mesurant la prise de conscience de la situation (SA) individuelle ou en équipe. Cette information est essentielle à la sélection d'une métrique appropriée afin de mesurer la SA dans une situation donnée tout en étant conscient des limites de cette métrique. Ce document est basé sur un relevé littéraire exhaustif de la mesure de SA et sur une recherche d'outils et méthodes utilisés pour mesurer ce concept. L'accent est mis sur les propriétés psychométriques des outils telles que la fidélité, la validité et la sensibilité. La revue de littérature révèle une grande variété d'outils, techniques et métriques, mais peu d'information concernant leurs propriétés psychométriques. Dans plusieurs cas, les données concernant la validité et la fidélité sont plutôt non concluantes, incomplètes et basées sur des échantillons restreints. Cependant, il semble y avoir des raisons pratiques expliquant le manque de données psychométriques : le faible nombre d'experts d'une situation donnée (c.-à-d. contrôleurs aériens, pilotes de jets et opérateurs dans des environnements de C2); et la complexité et les coûts élevés associés à l'évaluation avec simulateurs ou dans les environnements réels de travail (c.-à-d. cockpits, centres de C2).

Un deuxième objectif consiste en l'évaluation de la possibilité d'application d'outils de mesure de SA répertoriés dans la littérature aux différents environnements de C2. L'un des problèmes liés à la mesure concerne la grande variété de mesures pour un grand nombre de situations. On peut presque associer une technique différente de mesure à chaque situation potentielle dans laquelle un opérateur est impliqué. Chacune de ces situations possède son propre ensemble de procédures, de requis et de demandes liés à la tâche. Selon la répartition des outils de mesure de SA parmi les différents environnements de C2, le concept de SA peut être mesuré dans toutes ces situations. Dans certaines circonstances, une légère modification à une mesure peut être apportée afin de l'adapter à une autre situation.

Enfin, un autre problème concerne le manque de consensus parmi les chercheurs et les spécialistes en facteurs humains sur la définition et la modélisation de SA. Ainsi, on ne peut trouver un construit bien établi et universel sur lequel une évaluation peut être développée et validée. Néanmoins, on peut retrouver des mesures de SA bien établies telles que SAGAT et SART qui ont démontré leur utilité en relation avec l'évaluation ou le design de nouvelles interfaces et le développement de nouveaux programmes d'entraînement.

Breton, R, Tremblay, S., and Banbury, S. 2006. Measurement of individual and team situation awareness: A critical evaluation of the available metrics and tools and their applicability to Command and Control environments. RDDC Valcartier TR 2003-283. R&D pour la Defense Canada - Valcartier.

# Table of contents

---

Abstract.....	i
Résumé .....	i
Executive summary.....	iii
Sommaire .....	iv
Table of contents.....	v
List of figures.....	x
List of tables .....	x
1. Introduction.....	1
1.1 The Command and Control Process .....	1
1.2 The Command and Control Environments .....	3
2. Situation Awareness and its Measurement .....	5
2.1 Definitions of Situation Awareness.....	5
2.2 Individual Situation Awareness.....	5
2.3 Team Situation Awareness .....	6
2.3.1 Identification of team processes supporting TSA .....	9
2.3.1.1 Communication.....	10
2.3.1.2 Coordination of overlapping tasks .....	10
2.3.1.3 Information sources and types in SA and TSA.....	10
2.4 Measuring Situation Awareness .....	12
2.4.1 Measuring Individual Situation Awareness .....	12
2.4.2 Measuring Team Situation Awareness.....	13
2.5 Categories of SA Measurement Tools and Metrics.....	15
2.5.1 Selecting a Tool .....	15
2.6 Psychometric: Reliability, Validity and Sensitivity .....	15
2.6.1 Reliability.....	15
2.6.2 Validity .....	16

2.6.3 Sensitivity .....	16
3. Evaluation of the SA Measurement Tools .....	17
3.1 Measures of Eye Activity .....	17
Description .....	17
Psychometric Evaluation .....	17
Applicability to C2 Environments .....	18
3.2 Electroencephalographic Activity (EEG) .....	19
Description .....	19
Psychometric Evaluation .....	19
3.3 WOMBAT Situational Awareness and Stress Tolerance Test Battery .....	21
Description .....	21
Psychometric Evaluation .....	21
Applicability to C2 Environments .....	22
3.4 The FORCE_to_ASK technique .....	23
Description .....	23
Psychometric Evaluation .....	23
Applicability to C2 Environments .....	23
3.5 Situation Awareness Behaviourally Anchored Rating Scale (SABARS) .....	25
Description .....	25
Psychometric Evaluation .....	25
Applicability to C2 Environments .....	25
3.6 Think Aloud technique .....	27
Description .....	27
Psychometric Evaluation .....	27
Applicability to C2 Environments .....	27
3.7 Situation Awareness Global Assessment Technique (SAGAT) .....	29
Description .....	29
Psychometric Evaluation .....	29
Applicability to C2 Environments .....	30
3.8 SAGAT for Military Operations in Urbanized Terrain (-MOUT) .....	31
Description .....	31
Psychometric Evaluation .....	31

Applicability to C2 Environments .....	31
3.9 SAGAT Partial Blanking technique .....	33
Description .....	33
Psychometric Evaluation .....	33
Applicability to C2 Environments .....	33
3.10 Reaction Time to On-Line Probes (RT on-line).....	35
Description .....	35
Psychometric Evaluation .....	35
Applicability to C2 Environments .....	36
3.11 Situation Present Assessment Method (SPAM).....	37
Description .....	37
Psychometric Evaluation .....	37
Applicability to C2 Environments .....	38
3.12 Situation Awareness Linked Indicators Adapted to Novel Tasks (SALIENT).....	39
Description .....	39
Psychometric Evaluation .....	39
Applicability to C2 Environments .....	40
3.13 Situation Awareness Rating Scale (SARS) .....	41
Description .....	41
Psychometric Evaluation .....	41
Applicability to C2 Environments .....	41
3.14 Participant Situation Awareness Questionnaire (PSAQ).....	43
Description .....	43
Psychometric Evaluation .....	43
Applicability to C2 Environments .....	43
3.15 Situation Awareness Rating Technique (SART).....	45
Description .....	45
Psychometric Evaluation .....	45
Applicability to C2 Environments .....	46
3.16 Situation Awareness Subjective Workload Dominance Technique (SA-SWORD) .....	47
Description .....	47

Psychometric Evaluation .....	47
Applicability to C2 Environments .....	47
3.17    Map Reconstruction .....	49
Description .....	49
Psychometric Evaluation .....	49
Applicability to C2 Environments .....	49
3.18    China Lake Situation Awareness (CLSA).....	51
Description .....	51
Psychometric Evaluation .....	51
Applicability to C2 Environments .....	51
3.19    IMPLICIT Situation Awareness toolkit (IMPSA).....	53
Description .....	53
Psychometric Evaluation .....	53
Applicability to C2 Environments .....	53
3.20    Situational Awareness Supervisory Rating Form (SASRF).....	55
Description .....	55
Psychometric Evaluation .....	55
Applicability to C2 Environments .....	55
4. Evaluation of the Team SA Measurement Tools .....	57
4.1 DUO WOMBAT Crew Resource Management Test .....	57
Description .....	57
Psychometric Evaluation .....	57
Applicability to C2 Environments .....	58
4.2 Crew Awareness Rating Scale (CARS).....	59
Description .....	59
Psychometric Evaluation .....	59
Applicability to C2 Environments .....	59
4.3 Crew Situational Awareness (CSA) .....	61
Description .....	61
Psychometric Evaluation .....	61
Applicability to C2 Environments .....	61
4.4 Integrative Methods Approach .....	63

Description .....	63
Psychometric Evaluation .....	63
4.5 Targeted Acceptable Responses to Generated Events or Tasks (TARGETs) .....	65
Description .....	65
Psychometric Evaluation .....	65
Applicability to C2 Environments .....	65
4.6 Situation Awareness and Workload In-flight Measure (SAW-IM).....	67
Description .....	67
Psychometric Evaluation .....	67
Applicability to C2 Environments .....	68
4.7 Factors Affecting Situation Awareness (FASA) .....	69
Description .....	69
Psychometric Evaluation .....	70
Applicability to C2 Environments .....	70
4.8 Information Propagation in Team Environments (IPTE) .....	71
Description .....	71
Psychometric Evaluation .....	72
Applicability to C2 Environments .....	72
5. The Distribution of the SA Measurement Tools among C2 Environments .....	73
6. Applicability of Individual Measures of SA to TSA .....	77
7. Conclusion .....	79
References.....	81
List of symbols/abbreviations/acronyms/initialisms.....	91
Distribution list .....	93

## List of figures

---

Figure 1. The OODA loop .....	2
-------------------------------	---

## List of tables

---

Table 1: Cognitive processes listed in the different TSA definitions in the literature....	9
Table 2: Distinctions between SA and TSA in terms of information sources and types.....	11
Table 3. Distribution of the SA measurement tools among the C2 environments.....	74
Table 4. Distribution of the specific SA measurement tools among the C2 environments. ....	75

# 1. Introduction

---

In Command and Control (C2) environments, support systems (SS) are essential to support the human in the execution of the situation assessment (SA) and decision-making (DM) tasks. Paradoxically, with these sophisticated and powerful SSs, C2 environments may become more and more complex. With the growing volume of information made available by the technology and the accelerated tempo of the battlefield, the human information processing capability quickly becomes overloaded. Consequently, in order to adequately support the human, the design of SASS and DMSS must be shaped by both environmental constraints and human needs.

The design of such systems must be based, in part, on the evaluation of the human performance. From this evaluation phase, human needs can be identified and introduced into design requirements. In addition, data on human performance may also be useful to assess the effectiveness of a given SS. Then, the development of experimental paradigms in which measurement techniques are used is essential for the SS design and development.

The abbreviation SA has often be used to denote either the process of being aware of a situation (situation assessment) or the state resulting from that process (situation awareness). Breton and Rousseau (2001) proposed a classification of SA definitions found in literature. According to this classification, a definition can be either general or specific and either process- or state-oriented. Process-oriented definitions tend to be more general and state-oriented definitions to be specific. Obviously, measures have been developed to evaluate the process and others to evaluate the state. In this paper we use the current abbreviation "SA" to denote this concept as both the process and the state.

In the next sub-sections we briefly describe the C2 process. We also provide a set of factors characterizing the environments in which this process takes place. Finally, we present a group of individual and team SA measures found in the literature and we assess their applicability to C2 environments. These measures are evaluated according to their applicability to simulation environments and field trials for the navy, land and air force environments.

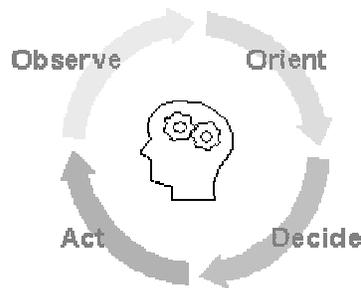
## 1.1 The Command and Control Process

Parks, Johnson, Picard, Dubois, MacLean, Ouellet, and Craig (1997), define C2 as "the process by which commanders can plan, direct, control and monitor any operation for which they are responsible". Pigeau and McCann (2000) consider C2 as "the exercise of authority and direction by a designated commander over assigned forces in the accomplishment of the force's mission".

Based on different models in the scientific literature defining C2, four general components can be identified. These components are:

1. *Observation*, that consists in gathering information from the situation;
2. *Situation assessment*, that includes data processing concerning the current situation;
3. *Course of action to execute*, decision-making and choice of a course of action;
4. *Implementation of the course of action* that was previously chosen.

These components are compatible with the OODA loop (Observe – Orient – Decide – Act), often used by the Canadian Navy (Paradis, Treurniet and Roy, 1998) to describe the C2 process. The OODA loop is illustrated in Figure 1. This loop represents taxonomy of general processes that lead to the implementation of the selected course of action.



**Figure 1.** *The OODA loop*

The OODA is a useful high-level representation of the basic processes in C2. But problems arise when the loop is viewed as a descriptive DM model, since it is limited by three basic difficulties. First, it has no representation of the feedback or feed-forward loops needed to effectively model dynamic DM. Second, it is a very high-level representation with abstract concepts; and third, it is a strictly sequential model with a single entry point. To address these deficiencies, other versions of the loop have been proposed. Fadok, Boyd and Warden (1995) proposed a loop with feed-forward and feedback arrows. While this gave to the OODA loop more dynamic properties, it was still defined with high-level concepts. In their version of the OODA loop, the M-OODA loop, Rousseau and Breton (2004) tried to provide more information on the nature of the actions executed within each module of the loop. It also included feedback and feed-forward loops to represent its dynamic nature. Further, in their conception of the loop, each module can serve as an entry point eliminating the bottom-up perspective of the classical version proposed by Boyd. Bryant (2003) provides a model of DM that is more consistent with modern theories of human cognition than the OODA loop. This model, called the CECA (Critique-Explore-Compare-Adapt), underscores the importance of mental models and top-down processing in the cycle.

Although they make substantial modifications to its structure, all these different versions maintain the same four general processes represented in the original OODA loop. In these different models, there is an information gathering phase coupled with

an understanding phase. These two phases are consistent with the SA model proposed by Endsley (1995b). Endsley asserts that decisions are formed by SA and SA is formed by decisions. However, Endsley underscores the importance of considering that SA is not DM, and DM is not SA. It is possible to have perfect awareness of the situation and still make non-optimal decisions. The opposite is also true; it is possible to make a sound decision (by luck) without thorough awareness of the situation.

The distinction between the SA and the DM processes is critical for the measurement of SA as a process and SA as a state. This measurement cannot be simply reduced to the measurement of DM performance. The process or state of SA must be measured independently of the quality of the decision. Consequently, an effective set of measures is essential for assessing either the process (situation assessment) or the state (situation awareness). The goal of SA measures is to define “what information is critical to the task, and how and when this information be presented?”

## 1.2 The Command and Control Environments

In C2 environments, the most problematic situations are ill defined and are characterized by a high degree of uncertainty. Moreover, situations may evolve rapidly and goals may shift constantly and compete with one another. Since these environments are dynamic, uncertain and ill defined, both the actions taken and results observed may be loosely coupled to one another, making it hard to distinguish effect from cause. Obviously, in most C2 environments, the stakes are high. They involve multiple players, doctrines and ROEs, as well as agency goals and standards, all of which can influence the selection of a potential course of actions.

The characteristics of C2 environments make human performance measurement a complex undertaking. Historically, laboratory settings and field trials have been used to measure human performance. Both settings offer advantages and disadvantages. On the one hand, laboratory settings provide an excellent control over the variables of interest. However, the level of control in laboratory environments makes the experimental setting too artificial, thus making the results hard to be generalized to C2 environments. On the other hand, although field trials provide a high level of realism, it may be difficult to control the extraneous variables that can jeopardize the validity of the experiment. These extraneous variables may alter the results observed and the conclusions drawn from the experiments. This situation has led to the development of powerful simulators that provide that provide a high degree of fidelity and realism and effective control over the experiment.

In the following sub-sections, we present a series of measures used to evaluate either the process or the end state of SA. For each measure we present a brief description and a critical evaluation of its psychometric properties. This evaluation is based on the available studies addressing the psychometric issues examined in the literature. Finally, we assess the applicability of each measure to C2 environments. Because of the characteristics of C2 environments, some measures may be better suited to evaluate the SA process (situation assessment) or the SA state (situation awareness).

This page intentionally left blank.

## 2. Situation Awareness and its Measurement

---

The term *situation awareness* has emerged from the aviation psychology community and has rapidly become a buzzword to represent an operator's understanding and knowledge of a complex system. It is widely believed that poor SA is a significant factor in aircraft accidents and incidents. The relationship between SA and human errors in such situations call for better understanding of the concept and the development of new ways to improve SA. Therefore valid and reliable techniques to measure SA are essential.

The present report is based on a literature survey on situation awareness measurement with regards to the psychometrics properties of the available metrics and tools. The aim of this document is to identify an SA metric for C2 environments. Before addressing the question of SA measurement psychometrics, the SA concept and its definitions are briefly discussed.

### 2.1 Definitions of Situation Awareness

The concept of situation awareness has been the subject of increased attention over the last decade (e.g., the 1995 special issue of *Human Factors*). There is, however, still some debate about the nature of the concept. The ongoing lack of precision in defining SA has made difficult to establish a coherent framework for conducting SA research. SA is intimately concerned with task performance during the operation of complex systems, such as aircraft, chemical, and nuclear power plants. It therefore seems logical to consider whether SA researchers can learn from other areas of research that have been concerned with task performance and behavior in the same domains.

### 2.2 Individual Situation Awareness

Knowledge of the information relevant to efficient task performance is critical to safety and productivity in a wide variety of situations. Indeed SA is applicable to all complex system operators such as air-traffic controllers, jet pilots, nuclear power plant operators and military commanders. Previous research that focused on SA has contributed primarily to our understanding of how individuals acquire and maintain SA.

Situational Awareness has been defined as a three-step process involving; 1) detecting or perceiving elements in the environment, 2) processing or comprehending the current situation, and 3) acting on the information or projecting the future status of the situation (Endsley, 1995b). There are other definitions that deviate more or less from Endsley's widely-cited definition.

SA is viewed by many to involve the cognitive processes underlying control of a dynamic environment (for a critical review of SA definitions and frameworks see Breton and Rousseau, 2001; see also Durso and Gronlund, 1999). SA is often

considered a ‘buzzword’ that serves as a label for a range of cognitive processes (see Prince, Salas, and Brannick, 1999). This view implies that SA should be conceptualized as a process (e.g., Sarter and Woods, 1995) rather than a state or product (Endsley, 1995b; as measured by most SA measurement tools). Some authors suggest that it is almost impossible to separate the two (Adams, Tenney, and Pew, 1995).

Nevertheless, it is important to make the distinction between SA as a process (assessment) and SA as a product (awareness). The product is the knowledge that a person has in a given situation. This knowledge can be measured with knowledge elicitation activities. The process is the set of activities that produce the knowledge. Measuring the process may be more complex since it may require indirect measure, through inferences, of cognitive processes. According to Sarter and Woods (1991), adequate situation assessments result in knowledge that may become part of situation awareness. Then, there should be significant differences between measuring the process and the result of this process.

SA process is sometimes treated as a skill given that novices can be trained to increase their SA. Obviously, considering SA process as a skill has repercussions on the way it is measured. Endsley and Garland (2000) identify many factors including meta-cognitive skills that can determine the level of quality of a performer. They conclude that training programs must address issues such as task management, comprehension, projection and planning and information seeking and self-checking activities. According to Shebilske, Goettl and Garland (2000), the definition proposed by Endsley (1995b) suggests that SA is a phenomenon mediated by controlled cognitive processes as opposed to automatic ones. Skill acquisition is viewed as a process in which increasingly sophisticated conditions are encoded into production rules, allowing experts in a domain to respond more quickly and accurately to situations they have previously encountered (Horswill and McKenna, 2004). This leads to a paradox. As an operator becomes more and more skilled, its SA process becomes automatic, unconscious and effortless. As a result, when interviewed, the operators might provide general, standardized and vague SA content. Its knowledge related to the situation, provided by automatic, unconscious and effortless process can be tacit or implicit. Consequently, SA process and product of skilled operators should be more difficult to measure.

## **2.3 Team Situation Awareness**

According to Dickinson and McIntyre (1997), a team consists of several individuals working together towards a shared goal. In order to achieve this, the individual team members must coordinate their work with other team members, so that the relevant information is shared. Efforts to improve teamwork must therefore focus on individual performance (Dickinson and McIntyre). At the same time, team members are dependent on each other to provide information and for coordination purposes. Those behaviours of team members that engender a sharing of information and a coordination of activities are collectively called teamwork (Dickinson and McIntyre). Brannick and Prince (1997, p.4), define a team as two or more people with different tasks who work

together adaptively to achieve specified and shared goals. There are many other definitions of team (Brannick, Salas and Prince, 1997), but they all seem to agree that teamwork is about working toward a shared goal and the definitions usually includes communication and coordination activities.

The importance of teams in the workforce has been well documented. For example, the success of military operations depends heavily on the effective performance of teams. In aviation, whether it is in the air traffic control tower or in the aircraft cockpit, safety critical tasks are often performed in teams where controllers or pilot crew can help each other and share the workload.

While at the earliest stages of the development of SA models and definitions, the subjects of interest of the research community were related to individual SA, defining and modelling SA has recently evolved to take into account the emerging concept of team cognition (Rousseau, Tremblay and Breton, 2004). Gutwin and Greenberg (2004) argue that awareness of other group members is a critical building block in the construct of team cognition, and consequently that computational support for awareness in groupware systems is crucial for supporting team cognition in distributed groups.

Another purpose of this review is, to examine the processes and behaviours that are associated with the concept of Team Situation Awareness (TSA). A breakdown in a team's process of acquiring and maintaining SA often leads to disastrous consequences. In fact, studies of accidents among major airlines attribute 88% of those mishaps to human error related to a problem of SA (Endsley, 1995a). TSA plays a critical role in aviation and military team decision-making, indeed many accidents and incidents have been attributed to failures in SA in team environments (Hartel, Smith and Prince, 1991; Endsley, 1988a). A number of authors have postulated that TSA can also be applied to other types of teams such as medical emergency teams and fire-fighters (Endsley, 1995a; Salas, Prince, Baker and Shrestha, 1995).

Such evidence reflects the importance to better understand the components necessary to improve TSA. Team tasks that are complex, dynamic, and information rich, often require that every team member obtain a certain level of SA based on the assessment of cues and events present in the environment. In turn, each team member's SA is modified as information is exchanged between members who may have observed an event that could be vital for the effective performance of the team. Attaining good TSA is a complicated process given that there are a number of interactive behaviors (e.g., information sharing, coordination) that play a crucial role in the achievement and maintenance of TSA.

In fact, a review of literature on TSA revealed that at least two components are crucial for achieving good SA: 1) individual SA, and 2) team processes that help build the SA of a team (Muniz, Salas, Stout, and Bowers, 1999). TSA and performance in team tasks can be significantly enhanced by the development of shared displays that are based on the shared information requirements of team members (Bolstad and Endsley, 2000). Most explanations of SA have focused on individual SA and have not been concerned with determining what is necessary for TSA. However, there is more to

TSA than simply combining the SA of individual team members (Schwartz, 1990). For example, Endsley (1995b) argues that whereas individual SA relies mostly on cognitive processes (i.e., perception, comprehension, and projection) TSA involves other activities, such as coordination and information sharing.

As it is the case for the distinction between SA product and SA process, there is a need for a clear distinction between SA and TSA definitions. The components related to each construct must be clearly defined. Next, we provide next a list of TSA definitions found in the literature.

**Bolman (1979)** refers to TSA as the crew's theory of the situation. Bolman viewed several behaviours as essential in developing, maintaining, and modifying a team's situation awareness. They include monitoring position-specific information, confirming and cross-checking information within the team, communicating relevant situation information to others, and coordinating activities.

**Schwartz (1990)** defines aircrew TSA as the accurate perception of variables that affect the aircraft and crew during a defined period. Schwartz states that each individual crewmember possesses a level of situation awareness, but that TSA could not be calculated simply as the sum of the situation awareness achieved by each crewmember. Schwartz argues that TSA is moderated by the pilot in command, who must receive information about the situation from each crewmember. Furthermore, Schwartz states that the level of situation awareness achieved is related to the level and quality of communication observed in the crew. Incomplete communication can be seen as an indicator of decreased situation awareness.

**Wagner and Simon (1990; as cited in Shrestha, Prince, Baker, and Salas, 1995)** define aviation TSA as the crew's understanding of flight factors that affect (or could affect) the crew and the aircrew at any given time and that subsequently have an impact on overall mission performance. Like Bolman (1979), Wagner and Simon suggest that aviation teams must monitor, process, and exchange information (i.e., mission objectives, orientation in space, environmental conditions, support, equipment status, and personal capabilities status) from several sources to maintain situation awareness.

**Prince and Salas (1993)** argue that each crewmember must seek and communicate information from both the internal and external environments. By communicating relevant situation information, crewmembers demonstrate knowledge of their overall mission goals and their individual task responsibilities. Furthermore, they assert that this information exchange among team members contributes to coordinated activity on the part of the crew.

**Wellens (1993)** defines TSA as "the sharing of a common perspective between two or more individuals regarding current environmental events, their meaning, and projected future status" (p. 272). He suggests that group SA could be maximized by having each member monitor different segments of the environment with enough overlap among members to ensure opportunities for coordination.

**Endsley (1995b)** suggests that TSA consists of both the situation awareness required of each team member and the overlap in SA that is necessary among team members, particularly for coordination. It is the “degree to which every crew member possesses the SA required for his or her responsibilities” (p. 39). Endsley’s concept of TSA is an add-on to her more established and formal concept definition of SA. She describes TSA consisting of the individual SA for each team member, plus the SA for potentially overlapping tasks (Kaber and Endsley, 1998). However, Hauland (2002) argues that her conceptualisation of TSA suffers from an ill-defined team concept, and does not address interpersonal relations and the problem of units of analyses for team. One reason for this may be the conceptual limitations of the Information Processing (IP) framework. Hauland argues that cognitive psychology in general, and IP frameworks in particular, have not traditionally addressed the situation context and team related variables. He also argues that Endsley does not elaborate on the concept of team in relation to SA. Indeed, Endsley describes a team simply as “several people that are working together”. Furthermore, she describes TSA as the degree to which every team member has sufficient SA regarding situation elements they are individually responsible for. Thus overall TSA is good if every team member has good individual SA. In addition, and independent from the overall TSA, comes the potential overlap of SA requirements between team members. These subsets of requirements constitute the need for team co-ordination (Kaber and Endsley, 1998).

### 2.3.1 Identification of team processes supporting TSA

The definitions presented above contain different processes required for the development of TSA. The processes can be gathered into two different categories, communication and coordination. These two elements, part of the NATO IST-019, TG006 list on team functioning elements, are often cited in the literature as being two important factors in teamwork. Table 1 presents, under the communication and coordination labels, the different team processes listed in the definitions above.

**Table 1: Cognitive processes listed in the different TSA definitions in the literature.**

Communication	Coordination
Confirming (Bolman, 1979)	Monitoring (Bolman, 1979; Wagner and Simon, 1990, Wellens, 1993)
Cross-Checking information (Bolman, 1979)	Not defined overlapping tasks (Endsley, 1995b)
Information exchange (Wagner and Simon, 1990; Prince and Salas, 1993)	
Sharing (Wellens, 1993)	
Seek (Prince and Salas, 1993)	

### **2.3.1.1 Communication**

To provide a basis for building TSA, team members need to have information that will help each individual develop relevant expectations about the entire team task. TSA depends on communication at several levels (Bolman, 1979; Prince and Salas, 1989, 1993; Schwartz, 1990). The process of perceiving environmental information is affected by expectations developed from the communication of knowledge about mission objectives, own tasks, other relevant tasks, team capabilities, and other factors associated with team performance (Prince and Salas, 1989). Then, as information is integrated and comprehended, interpretations provided by other crewmembers may affect that comprehension. Thus, as new information is perceived from the environment by individual team members and is collected and shared, the situation awareness of other team members may be modified accordingly (Bolman, 1979). Thus, any team processes that facilitate communication (e.g., assertiveness, planning, leadership that encourages open discussions) will help to build TSA.

Indeed, communication is a major component of teamwork process. Communication is the teamwork component that links the other components (Dickinson and McIntyre, 1997). Communication is the link between monitoring other team members' performance and providing feedback about that performance (McIntyre, Salas, Morgan, and Glickman, 1989). The term communication is often understood as verbal communication. However, communication can be conceptualised in terms of an explicit or implicit means of distributing information within and between teams.

### **2.3.1.2 Coordination of overlapping tasks**

Coordination (i.e. mutual adjustments between team members) is a central feature of teamwork (Brannick and Prince, 1997). This can be achieved using a combination of communication and monitoring. Monitoring refers to the observation of activities of other team members (Dickinson and McIntyre, 1997, p. 22). This implies that individuals, before they can be part of a team, must have the knowledge and skills to perform their individual tasks and have an understanding of the tasks of other team members (Cooper, Shiflett, Korotin, and Fleishman, 1984; Glanzer, Glaser, and Klaus, 1956; Larson and LaFasto, 1989). Monitoring team performance is a crucial component of teamwork (Cooper et al., 1984; McIntyre et al., 1989, 1990; In: Dickinson and McIntyre, 1997, p. 21-22).

### **2.3.1.3 Information sources and types in SA and TSA**

Following the definitions analysis presented above, TSA can be defined as the result of the cognitive processes sustaining individual SA and a set of activities that make the individual working as a team. These activities are gathered under the communication and coordination labels of Table 1.

In order to understand TSA, it is important to identify the sources and types of information provided by the SA activities in the team environment. In individual SA, the source of information is the environment and the information types are related to

the situation (data, facts, events, etc.). In TSA, the main information source is the still the situation with the addition of another one, the team members. Obviously, one can argue that the team members are, in fact, part of the situation, but for a matter of definition, it is useful to represent other team members as distinct information sources.

Team members are the source of two different types of information. A first source can be labelled “communication information”. This source provides information on other team members’ knowledge, opinion, advices, or thoughts. This source can be used to complete or validate the content of the individual SA. Activities listed in Table 1 under the communication label can provide this type of information. A second source of information, labelled “coordination information” concerns the information related to the execution status of the task by other team members (Who is doing what?), and the location of different sources of information. The effectiveness of a team should be dependent of the capacity of the team members to locate in the other team members, the required source of information (Who knows what?). Monitoring activities should be useful to provide these two types of information. Table 2 summarizes the differences between SA and TSA in terms of information sources and information types.

**Table 2: Distinctions between SA and TSA in terms of information sources and types.**

	Information sources	Information types
SA	Situation	<ul style="list-style-type: none"> <li>▪ Data</li> <li>▪ Facts,</li> <li>▪ Events</li> </ul>
TSA	Situation	<ul style="list-style-type: none"> <li>▪ Data</li> <li>▪ Facts,</li> <li>▪ Events</li> </ul>
	Team members	Communication: knowledge, thoughts, opinions, advices Coordination: Task execution status, information location

The distinction between SA and TSA is critical for measuring the concepts. While, measure of SA must provide tools to evaluate either the process (situation assessment) or the product (situation awareness), the measures of TSA must also provide means to evaluate the team capacity to develop a shared SA through the communication and coordination activities. The effectiveness of a team should be related to the capacity of team members to share relevant knowledge, to provide advices and opinions at an appropriate moment in the situation, to the capacity of a given team member to

monitor other team members work and finally, to its capacity to locate rapidly who, within the team, should provide a required source of information.

## **2.4 Measuring Situation Awareness**

### **2.4.1 Measuring Individual Situation Awareness**

Sarter and Woods (1995, p. 16) proposed that: “the term situation awareness should be viewed just as a label for a variety of cognitive processing activities that are critical to dynamic, event-driven, and multi-task fields of practice”. Such a point of view enables applied work on SA to proceed but in the long run can be detrimental to the field and foremost to the development of general SA measurement tools. The acceptance of a more precise and universal definition of SA would bring considerable advantages to the field and bring focus to research efforts. At present the concept of SA is rather fragmented, along context-specific environments (e.g., infantry, aviation, C2).

Progress in the development of SA measures has been hindered by the lack of consensus among researchers and human factor specialists on a single operational definition of SA. There is no well-established construct upon which a measure can be developed and validated. Nevertheless, there is a large number of measures that have been developed and applied to the assessment of SA. Most of these have been attempts to measure SA directly, through subjective measures (based on rating scales). A smaller number have addressed process, behaviour and performance indices of SA. Each type of SA measures has faults that make results questionable: for example, measures based on self rating are confounded with knowledge of the situation outcome (good or poor performance), measures taken through freeze techniques can be very intrusive and measures that require retrospective or recall techniques rely too heavily on memory.

Taxonomies proposed to classify the diversity of measures differ among authors. Some propose a division into six categories: retrospective, concurrent, subjective, process, performance, and signal detection (e.g., Gravell and Schopper, 1994). Others suggest three-category classifications: performance, subjective ratings and simulation (Gawron, 2000) or again subjective measures, query methods and implicit performance (Durso and Gronlund, 1999). One can easily get confused in reviewing those classifications; for example, SAGAT, perhaps the most commonly used measure of SA, is sometimes labelled as a performance measure (e.g., Gawron, 2000), a query method (Durso and Gronlund, 1999) or a subjective measure (Croft, Banbury, Aymeric, and Dudfield, 2000). In the latter case, ‘subjective’ refers to the fact that the development of SAGAT queries are based on someone’s (expert) judgement, whereas ‘subjective’ often refers to the mere assignation of a numerical value to the quality of SA (Endsley, 2000a). The inconsistency in classifying SA measures is due to the lack of consensus on a formal definition of the concept and the key issues yet to be resolved.

## 2.4.2 Measuring Team Situation Awareness

Previously in this document, we have presented different definitions of TSA. As it is the case for individual SA, TSA may suffer from a lack of consensus over a general definition. There are a number of issues unique to the measurement of TSA. Salas et al. (1995) emphasized the need for defining the situation for teams when attempting to measure TSA. The situation determines the team members' task assignments, which may have an important effect on the specific TSA requirements. However, the situation side of the SA concept is different for TSA, insofar as operators share the tasks related to information acquisition, anticipation of situation developments and the distribution of this knowledge within and between teams. This sharing of tasks requires coordination. The coordination of tasks requires certain team competencies, some of which are held at the level of individuals, and some of which cannot be reduced to individual competencies. Indeed, the mutual monitoring of each other enables the team members to take each other's perspectives, eventually resulting in the coordinated activities, so that the team will act like one coherent unit. Contrarily to TSA, such competencies are by definition not a requirement for individual SA. As a consequence, measuring TSA becomes rather complex, because it requires several units of analysis, not limited to the sum of measures of each individual's SA.

One way to address the complexity issue is to provide answer to three basic questions: 1) "What to measure?" 2) "When to measure?" and finally 3) "How to measure?"

**What to measure?** The measurement of a concept as immature as TSA will be problematic until there is a clearer understanding of what the concept represents. Two critical measurements should first be made for TSA: (1) individual SA and (2) the team processes that team members use to build and exchange information and enhance team coordination.

Given that shared mental models have been hypothesized to be an important component of TSA (Robertson and Endsley, 1994; Stout et al., 1994; Wellens, 1993), the compatibility of mental models among team members should also be measured. Although the measurement of TSA may prove to be difficult, the measurement of mental models may present an even greater challenge. A number of techniques are available to measure team member mental models (e.g., Schvaneveldt, 1990). Because the measurement of cognitive states is vital to the understanding of SA, measurement tools of the cognitive factors underlying the acquisition and maintenance of SA are needed (see FASA, this review). However, until such tools are more widely available, it is possible to gain insight about team SA by concentrating on individual and team situation assessment processes (Salas et al., 1995). By focusing attention on these more behavioural aspects of performance, it is possible to draw inferences regarding the nature and quality of TSA.

For example, some of the behaviours that have been suggested as being important to TSA are confirming and cross-checking information in the team, coordinating activities (Schwartz, 1990), sharing information (Wellens, 1993), planning, allocating tasks, and conducting pre-task briefings (Robertson and Endsley, 1994). Prince and

Salas (1989) identified behavioural indicators of TSA that can be observed and documented in crew communications when a team is performing. They include identifying a problem or potential problem, recognizing the need for action, attempting to determine the cause of discrepant information, providing information to another team member before it is required, noting deviations, and demonstrating an awareness of the task status and of one's own performance.

One main problem with team-related measures is the decision regarding the unit of analysis (Cannon-Bowers and Salas, 1997). The unit of analysis could be identified simply as the team, but teams do not do anything independent of individuals. Behaviourally defined response indicators will therefore relate to individuals also.

The team versus individual dichotomy is not a straightforward one (Cannon-Bowers and Salas, 1997). Certain team competencies may exist at the individual level; specifically, certain individual tasks may be carried out as a consequence of the team context. For example, the monitoring of colleagues' tasks would constitute individual taskwork, but it is at the same time clearly an aspect of teamwork (taking perspectives, coordinating). According to Cannon-Bowers and Salas (1997), team competencies exist at different levels of measurement, (i.e. with corresponding units of analysis):

(1) Individual Competencies (Individual Level): Knowledge, skills and attitudes required to meet the individual job requirements.

(2) Team Competencies held at the Individual Level: These competencies are generic to all teams. For example, the individual monitoring of colleagues' tasks is carried out independent of the colleagues, although such monitoring requires their presence. Thus, it is considered a team component, measured at the individual level, that the team members monitor each other's task and performance on those tasks.

(3) Team competencies held at the Team Level: These competencies constitute unique inter-personal relations in a team. For example, the use of explicit (i.e. verbal) or implicit (i.e. non-verbal gestures) communication between two team members may depend on the team members' common professional (and personal) history in the team. Such relations cannot be meaningfully reduced to a set of individual competencies. Thus, behavioural measurements at the team level will therefore consist of relations between two (or more) sets of individual behaviors. Direct measures of teamwork, such as questionnaires, can have this relation built into the questions (e.g. to what extent did the team members switch roles), or the measure could be computed based on the comparison of two independent answers (e.g., mutual awareness). The uniqueness of such relations may be questioned, possibly adding an additional category to this measuring level (i.e. generic or unique).

**When to measure?** Team SA is not a static state but the result of recurrent processes (i.e., information seeking, information processing, and information sharing) that take place within a team. Therefore, a single measure is not adequate for measuring TSA throughout a particular task. To produce higher temporal stability for both situation assessment and situation awareness, measurements should be made over a series of key events while the team is performing its tasks (Salas et al., 1995).

**How to measure?** Sarter and Woods (1991) suggested the use of complex scenarios to measure TSA. They suggest an approach, which involves the embedding of events within the scenarios to elicit key situation assessment behaviours and processes. They also recommended that a number of such events be embedded in order to provide multiple opportunities to measure TSA. Methods that allow observers to document and rate TSA throughout the scenario will need to be developed. For example, checklists, behavioural-anchored rating scales, and behavioural observation scales (for a review see Flin and Martin, 2001) have been used to evaluate individual and team performance. These formats have also been employed in team process research with some success (Baker, Salas, Cannon-Bowers, and Spector, 1992; Brannick et al., 1995; Glickman et al., 1987; Morgan et al., 1986).

## **2.5 Categories of SA Measurement Tools and Metrics**

The classification adopted in the present report suggests two orthogonal dimensions: on-line vs. off-line crossed with direct vs. indirect (Hoffman, 1997). The first dimension is concerned with the moment at which the measure is taken (concurrent vs retrospective) and the second one refers to whether the measure is inferred from behaviours or based on the operator's perception.

### **2.5.1 Selecting a Tool**

The selection of an appropriate SA measurement tool can be a difficult task given the wide variety of measures, each one being appropriate for a slightly different purpose. Moreover, all measures have their benefits and costs; there is no single measure that is better, for all purposes and concerns, than the others. Since most definitions of SA are product-based, most of the published measures are aimed at assessing the product of SA, while only a few are aimed at the traits, processes or conditions that underlie the development of SA. Beyond this distinction, the choice of measure is a function of the ease of implementation (thus favouring subjective measures), obtrusiveness, and psychometric attributes such as validity, reliability and sensitivity.

## **2.6 Psychometric: Reliability, Validity and Sensitivity**

### **2.6.1 Reliability**

Information about the test's reliability (i.e., how stable or consistent a measure the test is) and the strength of its validity (i.e., how well it actually measures what it claims to measure) is critical in the process of selecting a metric or tool. Validity depends on reliability and a test cannot be more valid than it is reliable. Reliability can be determined in a number of ways: i) Test-Retest: the same test to the same individual more than once. ii) Split-Half: the test is divided into two – one half is given at time 1 the other at time 2. This method also provides a measure of the internal consistency of a measure. Another type of reliability that is relevant to subjective rating measures is the inter rater reliability.

## 2.6.2 Validity

A measurement tool or technique must be valid. Does it measure what it purports to measure? There are four main types of validity:

- i) Content validity: it is essentially concerned with the operationalization against the relevant content domain for the construct. This approach assumes that there is a detailed description of the content domain, but for constructs such as SA (e.g., self-esteem, intelligence), it is difficult to decide on the criteria that constitute the content domain;
- ii) Convergent validity: Does the test correlate well with other established measures of SA? (e.g., self report with supervisor's assessment; SAGAT and SART);
- iii) Predictive validity: in the case of SA measurement, it is the correlation between level of SA and task or simulation performance;
- iv) Construct validity: refers to the extent to which a tool or technique measures the concept it is designed to measure. Do the questions or items of a metric relate to the theoretical construct it purports to measure?

With regards to ecological validity, any attempt to stage an event (simulation), or to introduce a confederate to draw out the operator's thinking, biases the operator's thought processes and attention, and artificially modifies the environment.

## 2.6.3 Sensitivity

Sensitivity, diagnosticity and selectivity are three other criteria of measurement tools and techniques that are found in the literature. Sensitivity is the ability of a measure to identify any changes due to the impact of different factors, or distinguish between different conditions (e.g., ageing). To most authors, diagnosticity refers to the degree to which a measure identifies the cause of any variation. The criterion of diagnosticity goes beyond the simple identification of changes, a measure with good diagnosticity assists in identifying why there were changes in SA (definition used in the present report). However, the distinction between sensitivity and diagnosticity is subtle and across reports the two are often used interchangeably. Moreover, to some authors diagnosticity refers to selectivity or again for some others it seems to refer to prediction of performance (predictive validity). Selectivity is the degree to which a measure is sensitive to changes specific to SA. In the context of SA measurement, depending on the construct a measure is based upon, it is preferable if the measure does not correlate too highly with measures of workload or memory span.

### 3. Evaluation of the SA Measurement Tools

---

In this section, we present a list of SA measurement tools. When known, developers are cited. The classification is based on the direct-indirect and online-offline distinction. Also, we identify the locus of measurement (concurrent versus post). Following a brief description of the measurement tool, a psychometric evaluation of the tool is provided as well as its applicability to C2 environments.

#### 3.1 Measures of Eye Activity

<b>Developers:</b>	-----
<b>Classification:</b>	Indirect / Online
<b>Locus of measurement:</b>	Concurrent

##### Description

There are several measures of eye activity: blink amplitude, blink duration, scan pattern, pupil dilation, glance duration, fixation duration and dwell time.

***Eye-Tracker: Eye Point of Gaze (EPOG).*** The EPOG is a technological system used to determine where the operator is looking or where he/she should look. While the point-of-gaze brings information about what the operators are looking at, the pupil dilatation brings information about the efforts elicited by the incoming stimuli or the activity performed at the time of gaze.

***Eye-Tracker: Blink Duration.*** Eyeblinks are recorded by a technological system that measures their frequencies. There is little information about the psychometric attributes of eye tracking (in general) in the context of SA. There is a validation study by Vidulich, Stratton, Crabtree, and Wilson (1994) that investigated eyeblinks in tasks that allowed either a poor SA or a better SA. In the poor SA condition, the participants expressed a higher rate of eyeblinks that were of a shorter duration.

These measures are indirect. The measure of SA is based on inferences. Consequently, they ask for a certain coupling between what they measure and the concept of SA. Assumptions are required to establish that coupling.

##### Psychometric Evaluation

***Sensitivity and Diagnosticity:*** Highly sensitive to changes in the environment or in the task. There is very little information available about the diagnosticity.

**Selectivity:** As most process measures, it is not specific to SA. In relation to the allocation of attention the assumption that looking is seeing (registering information) has to be made.

**Reliability:** Eye-trackers are highly reliable.

**Validity:** To our knowledge there was no information available at the time of writing this report.

## **Applicability to C2 Environments**

EPOG are providing data about the eye activity. It provides information about where the operator is looking. This type of physiological measure requires sophisticated data recording systems. Moreover, eye-trackers require displays and radar screens to determine where the operator is looking on a specific display at a given moment.

Because of the importance of the technological devices required to collect the data, even with the latest versions, this type of measure is hardly applicable to field trials. Most of eye-trackers ask for a certain level of stability of the operator's head as well as for the technological settings. They can also considerably limit the operator's mobility. One of the main reasons the do field trials is related to the importance of experimental setting realism. Since, the eye-trackers are not part of the real environment, the introduction of this device in real setting may reduce considerably the level of realism.

Eye-trackers are better suited to be coupled with simulators. They may be dissimulated into the simulator devices. For instance, they may be inserted into a pilot helmet. Because one of the objectives with the eye-tracker is to know where the operator is looking, they require displays. Then, eye-trackers may be appropriate in navy and air environments in which most of the information is obtained through displays. However, with the recent developments of virtual reality, it may be possible to use eye-trackers in the context of infantry operations. The Soldier Information Requirements Technology Demonstrator project (SIREQ\_TD) conducted at DRDC Toronto includes such experiments with eye-trackers in virtual reality environments.

Ideally, eye-trackers should be used in experimental settings including displays, consoles, radar screens, etc. The operator should be seated in front of those devices to allow a certain level of stability required by the eye-tracker equipments. New versions of eye-trackers are available. These versions are smaller and less obtrusive. However, they are fragile and still restrictive for the operator movements.

## 3.2 Electroencephalographic Activity (EEG)

<b>Developers:</b>	-----
<b>Classification:</b>	Indirect / Online
<b>Locus of measurement:</b>	Concurrent

### Description

The electro-encephalography (EEG) is a technique for monitoring brain activity. EEG is a non-invasive procedure that can measure biological activity through the skull and reveal the human brain at work. EEG uses electrodes placed on the scalp to detect and measure patterns of electrical activity emanating from the brain. Sensitive electrodes are attached to the head, and the signals are amplified to give a graph of electrical potential versus time. It can be measured and compared at different spots on the head simultaneously, to give a 2-dimensional activity map of the cerebral cortex. EEG can determine the relative strengths and positions of electrical activity in different brain regions. It is also the only way of measuring brain activity in real time, since the time aspect of the graph is measured in milliseconds. EEG can record complex patterns of neural activity occurring within fractions of a second after a stimulus has been administered. Brain waves can be recorded for the occurrence of a stimulus, and the experiment is repeated several times. The graphs are then averaged, and the resulting data are called event-related potentials (ERPs). ERPs have also been useful for finding out how long the brain uses to process different kinds of information, and monitor levels of attention and stress for various situations. By tracking changes in activity, one can determine brain areas and patterns of activity that mark these changes. It is possible to record the brain activity of people while they are performing a task or activity.

Here again, EEG is an indirect measure based on inferences. Consequently, they ask for a certain coupling between what they measure and the concept of SA. Assumptions are required to establish that coupling.

### Psychometric Evaluation

***Sensitivity and Diagnosticity:*** EEG are highly sensitive to changes in the environment or in the task. To our knowledge there was no information available for the diagnosticity at the time of writing this report.

***Selectivity:*** As most process measures it is not specific to SA. Again, there is little information available and results are inconclusive. Vidulich, Stratton, Crabtree, and Wilson (1994) investigated EEG in tasks that allowed either a poor SA or a better SA. In the poor SA condition, the EEG spectrum indicated higher theta frequencies at the expense of the alpha frequencies. Perhaps, these cues are more indicative of a difficulty with mental workload, or of the activities associated by the process of situation assessment.

**Reliability:** Variability (due to data reduction).

**Validity:** To our knowledge there was no information available at the time of writing this report.

### **Applicability to C2 Environments**

EEG provides information about the brain activity. EEG asks for important technological settings. It requires electrodes, wires and sophisticated data recording systems that may reduce considerably the mobility of the human subject.

Because of the importance of data recording system, the EEG can hardly be used in field trials. The introduction of this device in real setting may reduce considerably the level of realism that is essential in field trials experiments.

As for the eye-trackers, EEG is better suited to be coupled with simulators. They may be dissimulated into the simulator devices. For instance, the electrodes may be inserted into a pilot helmet. EEG can be used in experimental environments simulating navy, air and infantry operations (in virtual reality context).

EEG should be used in experimental settings allowing the wiring of operators with data acquisition systems. Consequently, the operator's mobility is considerably restricted. EEG can be used in situations where an operator is seated in front of displays and other technological devices.

### 3.3 WOMBAT Situational Awareness and Stress Tolerance Test Battery

<b>Developers:</b>	LaRoche and Roscoe
<b>Classification :</b>	Indirect / Online
<b>Locus of measurement:</b>	Not applicable

#### Description

The WOMBAT is a modern psychological assessment tool for selecting complex system operators such as pilots, air traffic controllers, ship operators, nuclear-plant operators, in fact anyone in charge of complex operations involving multiple concurrent inputs and response alternatives. WOMBAT may have the look and feel of a video game. However, it is much more than a complex game; it is a powerful, culture-independent PC-based set of tasks. These include target tracking, spatial orientation, pattern recognition and working memory. It is claimed by the developers to be the ideal system for measuring situational awareness under stress.

The procedure is simple. The applicant sits in front of a desktop computer and interacts with it using a special WOMBAT console. The console has two standard joysticks and a custom keypad. Performance measures have the advantage of being objective, usually non-invasive, and easy to collect.

Three versions of WOMBAT are available. The individual WOMBAT-CS (Complex Systems) measures the ability of a person to manage complex systems. The Duo WOMBAT-CS measures the same ability of a person while interacting with another operator. The WOMBAT-FC (Flow Controller) measures the ability of a person to keep track of a continuous flow of information in dynamic situations.

#### Psychometric Evaluation

**Sensitivity and diagnosticity:** Global, primary performance measures have the problem of lack of sensitivity and diagnosticity. If global performance is the only human/system measure important design differences may be masked, poor subtask performance may go undiagnosed, and the impact of SA may be melded with many variables and go unnoticed. Poor decision-making, the end process of a string of cognitive processes – and the resulting mission performance, if taken as the only measure of SA, may not disclose with any specificity what SA variables have been at work.

The WOMBAT test battery is sensitive to variation in SA. At the occasion of a gliding competition, the WOMBAT test battery discriminated between experienced pilots who are considered as the elite and those experienced ones who are not. The WOMBAT score was correlated with only one of four abilities that were hypothesised to underlie performance, namely pattern recognition, while short-term memory, visual search recognition, and spatial imagery were not (O'Hare, 1997).

**Selectivity:** O'Hare and O'Brien (2000) further tested if the WOMBAT score could be explained by other factors than situation awareness. Their results indicated that general intelligence could explain 28 percent of the variance, leaving the remaining 72 percent to other factors that cannot be age, experience with computer or video games, or the preceding underlying performance factors. Roscoe, Corl, and LaRoche (2001, p. 52) claimed that the remaining 72 percent is solely influenced by the situation awareness capability.

**Reliability:** After a period of practice, the performance reaches an asymptote. In consequence, the WOMBAT has an excellent test-retest liability.

**Validity:** The WOMBAT test battery is claimed to have a good predictive validity. O'Hare and O'Brien (2000) found that the WOMBAT score predicted early performance on TRACON, an air traffic control task that requires an excellent SA. When high levels of situational awareness are demonstrated on the test, individuals can be expected to perform well on complex jobs. However the latter argument is questionable; SA as measured by WOMBAT is inferred from performance. This is related to the problems of specificity mentioned above.

### **Applicability to C2 Environments**

WOMBAT is a test battery with its own experimental setting including displays, joysticks and consoles. The measurement of the SA concept with WOMBAT is an experiment itself with its own devices, protocols and settings. It can only identify, from its own experimental setting, which individuals have the best SA capability.

Because of the nature of WOMBAT, it can hardly be inserted into infantry, air and navy operations during field trials. The activities required by WOMBAT may alter considerably the realism of the trials. Also, the inclusion of WOMBAT test battery may not be appropriate within simulation environments. Consequently, it can hardly be coupled with experimental settings.

WOMBAT can be used before field trials or experiments with simulators to determine a priori the human potential to develop a good level of SA.

### 3.4 The FORCE\_to\_ASK technique

<b>Developers:</b>	Payne, Bettman and Johnson (1993)
<b>Classification:</b>	Indirect / Online
<b>Locus of measurement:</b>	Concurrent

#### Description

The technique consists of instructing participants that they have to use mouse-clicks in order to get the information they need to make appropriate decisions. The force-the-user-to-ask technique was used by Ricks, Jonsson, and Barry (1996) as a SA measure during a simulated aircraft flight. Andersen and Hauland (2000, p. 268) suggested that “the degree of correspondence between a pre-defined ideal behaviour and the observed behaviour may constitute a measure of situation awareness.” Then, with the force-the-user-to-ask technique, the inferences made from the behaviours observations can lead to the identification of a specific SA content or to the comparison with an achievable SA content inferred from the observation an achievable ideal behaviour.

This technique was developed within the context of decision-making and adapted to SA measurement by Ricks, Jonsson, and Barry (1996) and Andersen and Hauland (2000).

#### Psychometric Evaluation

To our knowledge there was no information available at the time of writing this report.

#### Applicability to C2 Environments

The observed behavior technique (asking a specific information by clicking with a mouse) must be seen as the request of specific information required by the execution of a given task.

Since it requires a mouse, this technique may not be used in field trials. Moreover, even if a mouse is already available in the real environment, the Force\_to\_Ask technique may drastically change the execution of a given task. The information-gathering task (by clicking) may be significantly different than the task executed in the real environment. Consequently, the use of this technique in real settings may reduce considerably the level of realism. The problem of realism is also true in simulation environments.

Obviously, the clicking technique may not be appropriate in virtual environments simulating infantry operations and flight simulators. This technique may be more appropriate in environments including displays, joysticks and mouse. In that sense, the operation rooms of frigates may be more adequate to use this technique.

This page intentionally left blank.

### 3.5 Situation Awareness Behaviourally Anchored Rating Scale (SABARS)

<b>Developers:</b>	Matthews, Pleban, Endsley and Strater (2000)
<b>Classification:</b>	Indirect / Online
<b>Locus of measurement:</b>	Concurrent

#### Description

*Observer rating (in general).* The observer rating technique delegates the impartial observer role to a peer of the person under study. The content of SA is inferred from the observations. Observer rating as an SA measurement technique involves recruiting subject matter experts, or knowledgeable observers to rate the SA of an operator, with the advantage of the observer having access to more (or less) precise performance criteria. A potential drawback with this technique lies in the inability of the observer to know what the operator's internal concept of the situation is, what variables the operator is potentially discarding in order to process an event, or what information-processing the operator is undertaking on the variables that are being accessed. The SABARS was specifically designed for infantry operations at the platoon level but it can also be adapted to other infantry levels. An expert observer has to check on a five-point scale the quality of 28 behaviours that indicate an appropriate SA. A validation study of SABARS as well as SAGAT-MOUT and PSAQ is underway (Matthews, Pleban, Endsley, and Strater, 2000).

#### Psychometric Evaluation

To our knowledge there was no information available at the time of writing this report.

#### Applicability to C2 Environments

The observer technique provides information about operator's behaviors. From these behaviors, the SA content can be identified. Then, the observer rating technique involves the inclusion of an observer (most of the time a peer) into the task environments. Consequently, the limitation of this technique is related to the possibility of inclusion of this person. This technique can be used if the inclusion of the observer does not reduce the level of realism of the experimental environments.

In field trials, the observer rating technique can be used in infantry operations (SABARS). The observer rating can also be adapted for the operation room of frigates. It may be adapted to be used in real flight situations if there is room for the observer. The observer rating is appropriate in simulation environments. A benefit of this technique is that the operator's task is not altered. Consequently, if the presence of the observer is possible and does not reduce the level of realism of the simulation, the observer rating technique can be adapted to most of the simulation environments.

This page intentionally left blank.

### 3.6 Think Aloud technique

<b>Developers:</b>	Based on the use of verbal protocols
<b>Classification:</b>	Direct / Online
<b>Locus of measurement:</b>	Concurrent

#### Description

Verbal protocols can be regarded as a form of the concurrent SA measure. Typically, operators are asked to ‘think-aloud’ as they perform a task. This reporting may take place in real-time, or later during a video replay of events. The technique does, however, suffer from many of the same problems as the concurrent probe technique, and there is some evidence that the process itself might alter the way in which the task is performed. Think aloud can be considered as a behaviour or process measure that attempts to determine an operator’s information acquisition and attention allocation. Verbal protocols represent a method for abstracting the process by which an operator develops or maintains SA (Sullivan and Blackman, 1991). This method has been criticized for the problems that come with knowledge elicitation such as the difficulty to extract knowledge from experts and the risk of post-hoc assessment of SA based on the reconstruction of events (see Ericsson and Simon, 1980; 1984). Pew (1995) suggests that this (concurrent verbal protocol) SA assessment technique is most useful early on in development, and while disruptive, may be worth the cost from a long-term developmental standpoint.

#### Psychometric Evaluation

To our knowledge there was no information available at the time of writing this report.

#### Applicability to C2 Environments

The Think Aloud technique provides information about operator’s logic under their behaviors. The operator must do the effort to verbalize. For this reason, this technique may interfere with the non-verbal aspects of the on-going processes. It does not necessarily require important data recording devices. Then, this method is not very obtrusive. The information can be recorded using a simple data recording device ( e.g., tape recorder) or by an observer.

The Think Aloud technique may be appropriate for field trials testing. However, since the operator must describe his behaviors concurrently with the execution of the actions, there is a high probability that the execution of task may be altered. In that sense, the use of this technique in field trials may reduce considerably the level of realism required by these experiments. Moreover, it may also require recording devices. However, the alteration of the studied task seems to be the only major cost related to this technique. It can be used in simulation environments.

This page intentionally left blank.

### 3.7 Situation Awareness Global Assessment Technique (SAGAT)

<b>Developer:</b>	Endsley (1987; 1988b)
<b>Classification:</b>	Direct / Online
<b>Locus of measurement:</b>	Concurrent

#### Description

SAGAT is not really a metric as such but rather it is the operational definition of SA as developed by Endsley and her collaborators. SAGAT provides users with a set of procedures to apply to a given simulated situation. SAGAT requires the experimental process (e.g., a simulation) to be suspended or frozen at some point, and the participant to be quizzed as to their assessment or knowledge of the environment or system state before the trial resumes. SAGAT is a complex method applied in a number of steps. SAGAT provides a snapshot of the operator's mental model, can cover a wide range of SA elements, provides a direct measure of the accuracy of the participants SA (self) assessment, and can be easily collected and assessed. Among its limitations SAGAT requires the simulation be frozen to collect data, is intrusive to the experimental task, and is at least partly performance driven (Endsley, 1995a). The measure of SA is calculated from correct responses to online queries about the situation. SAGAT is claimed to be an objective measure as it is not based on subjective ratings (response) but the fact that queries are developed based on a cognitive task analysis performed by subject matter experts makes the measure subjective in the sense that it is based on someone's opinion. SAGAT is perhaps the most common measure of SA and has been the object of several validation studies. However there is a different SAGAT for every situation and its application is restricted to simulations.

#### Psychometric Evaluation

**Sensitivity and Diagnosticity:** The sensitivity of the SAGAT has been recently reviewed by Vidulich (2000). SAGAT is sensitive in a diversity of application fields. The diagnosticity is claimed to be high.

**Selectivity:** Can be considered as high given that each measure based of SAGAT is specific to the simulation it was developed for. However, the attempts to combine a set of SA queries into one global SA score failed (Endsley, 1998).

**Reliability:** Test-retest coefficients are very high (see Endsley, 2000a).

**Validity:** Criterion validity of SAGAT has been reasonably tested (Endsley, 2000a). SAGAT is linked to performance (Endsley, 1990), to the situation complexity (Gugerty, 1997; Endsley and Rodgers, 1998), and to the need to prioritize an important aircraft (Gronlund, Ohrt, Dougherty, Perry, and Manning, 1998).

## **Applicability to C2 Environments**

The SAGAT technique is used to capture the individual SA content. There are two major requirements related to the application of this method. First, it must be possible to freeze the action at specific or random periods of time. Second, the queries must be based on a task analysis. During this freezing moment, the operator is queried about his actual knowledge of the situation (his SA content about the three levels). At this moment, the operator's screens and displays are blanked. The responses of the queries must rely on the memory content that the operator has on the situation.

The freezing requirement by the technique may alter considerably the progress of the scenario included in the experiment. Even if Endsley (1995a; 2000a) did not find an effect of the interruption of the human performance, this freezing technique is still seen as very obtrusive and reducing considerably the level of realism of the experimental settings. Also, it is not known for sure whether the operator returns to the situation as well as he was before the freezing nor whether he maintains the same mindset in regard with the task execution.

Because of the freezing requirement, this technique can hardly be used in field trials experiments. This technique is better suited to be used in simulation environments. The technique can be used in flight simulators and naval environments. The SAGAT-MOUT technique (see below) is an adaptation of the SAGAT technique to infantry operations.

### **3.8 SAGAT for Military Operations in Urbanized Terrain (-MOUT)**

**Developers:** Matthews, Pleban, Endsley and Sarter (2000)

**Classification:** Direct / Online

**Locus of measurement:** Concurrent

#### **Description**

Recently, Endsley and her colleagues (Endsley, Holder, Leibrecht Garland, Wampler, and Matthews, 2000; Matthews, Pleban, Endsley, and Strater, 2000; Strater, Endsley, Pleban, and Matthews, 2001) have concentrated on adapting SA measures for the infantry to support the individual soldier to the brigade level in order to inform information technology development. Matthews, et al. (2000) developed SAGAT queries specific to infantry operations in virtual MOUT. This work is still in its infancy and more research is required to develop and validate this measure within infantry operations.

A validation study of SAGAT-MOUT as well as SABARS and PSAQ (applied to infantry) is underway (Matthews, Pleban, Endsley and Strater, 2000).

#### **Psychometric Evaluation**

To our knowledge there was no information available at the time of writing this report.

#### **Applicability to C2 Environments**

The SAGAT – MOUT technique is used to capture the individual SA content in infantry operations in natural settings. This technique requires the freezing of the complete field exercises. During this freezing moment, the soldiers are queried about their actual knowledge of the situation (their SA content for the three SA levels).

If the freezing requirement can be met, this technique can be applied during field trials. However, as with the SAGAT technique, the application of the SAGAT-MOUT can be very obtrusive and may alter the task execution. Also, it is not known for sure whether the soldiers get involved in the situation as well as he was before the freezing. Nor whether they present the same mindset in regard with the task execution. Actually, there is no version of this technique used in simulation environments. However, it would be possible to use the SAGAT-MOUT technique in virtual environments.

This page intentionally left blank.

### 3.9 SAGAT Partial Blanking technique

<b>Developer:</b>	Hillburn (2000)
<b>Classification:</b>	Indirect / Online
<b>Locus of measurement:</b>	Concurrent

#### Description

Typically, during SAGAT queries, the action is frozen and the simulator screens are blanked. This blanking screen technique can be argued to restrict SA measurement to its sole short-term memory component. If an operator developed good comprehension (level 2 SA) of the situation, he might not remember an exact parameter, but he might know where to look for. When the screens are blanked, the operator cannot demonstrate this understanding. During the time screens are blanked, Endsley (1995a) observed only a slow decay of memory between 20 seconds and 6 minutes. This observation answered to concerns about a possible lack of short-term memory interference, but it does not answer the second concern that such a blanking weakens the comprehension assessment.

A more critical stance has been taken by Sarter and Woods (1991) who assert that, when using SAGAT, the operator's SA is not being assessed; rather, the pilots are simply answering the question of what information they could recall when prompted, whilst deprived of a dynamic environment. Hillburn (2000) proposed to use a partial blanking technique as a compromise. During a flight simulation, an air-traffic controller experiences the failure of a secondary radar (loss of all text) while all other radar functions remain. If the controller can answer to the SAGAT query immediately, then it is assumed that the information was part of the SA content. If the right answer follows a reference to the blanked screen, then the experimenter is cued about a level 2 SA (comprehension query). With this partial blanking technique, the SA content is inferred from the observation of specific behaviours related to SA, which makes this procedure indirect rather than the assumed direct measurement of SAGAT.

#### Psychometric Evaluation

To our knowledge there was no information available at the time of writing this report.

#### Applicability to C2 Environments

The SAGAT Partial Blanking technique is actually used to capture the pilot SA content in flight simulations. As with SAGAT and SAGAT – MOUT, the application of this technique requires the freezing of the experimental environment. During this freezing moment, the operator is queried about his actual knowledge of the situation (his SA content). In comparison with the two other SAGAT techniques, the operator's

screens and displays are partially blanked. However, it seems that this technique suffers from the same problem that the two others. The freezing moment may alter considerably the progress of the simulated scenario included in the experiment. Then, this freezing technique may be very obtrusive and may reduce considerably the level of realism of the experimental environment. Also, it is not known for sure whether the soldiers get involved in the situation as well as he was before the freezing. Nor whether they present the same mindset in regard with the task execution.

Because of the freezing requirement, this technique can hardly be used in field trial experiments. Actually, this technique is used with pilots in flight simulations. The technique asks for the blanking of displays and screens used by the operator to obtain the required information for the SA content. Consequently, displays and screens are indispensable to use the partial blanking technique. Then, it would be possible to develop a version of this technique adapted to the naval operations in the frigate operation rooms.

### 3.10 Reaction Time to On-Line Probes (RT on-line)

<b>Developers:</b>	-----
<b>Classification:</b>	Direct / Online
<b>Locus of measurement:</b>	Concurrent

#### Description

Reaction time to online probes is a technique that queries operators for specific aspects of a given situation, at all three levels of SA. This technique is derived from SAGAT but the procedure is different: probes are queries addressed to the person under study without stopping the ongoing activity and one probe at a time. This technique is suited when the pace of a situation is relatively slow or when there is some period of inactivity (Pew, 2000). The format of queries allows the users to employ time to respond rather than accuracy as SA measure.

The advantage of real-time probes over SAGAT is their applicability in situations that cannot be interrupted. The real-time probes are less intrusive than the SAGAT queries. It offers a good control over the measurement process without threatening the level of situation realism. However, fewer queries can be asked during a particular lap of time. Thus, this measurement is less exhaustive. Also, even though the probes do not interrupt the on-going action, they can nevertheless interfere with the cognitive processes. Up to now, the real-time probes are a promising new measurement of SA.

#### Psychometric Evaluation

**Sensitivity and Diagnosticity::** Real-time probes sensitivity is moderate but it appears to increase in sensitivity with the number of queries (Jones and Endsley, 2000a). Another study found good sensitivity for SAGAT but not for RT online probes (Endsley, Sollenberger, and Stein, 2000). Concerning the diagnosticity, very little information available and validation studies underway.

**Selectivity:** To our knowledge there was no information available at the time of writing this report. Validation studies underway.

**Reliability:** To our knowledge there was no information available at the time of writing this report. Validation studies underway.

**Validity:** Concurrent validity: low correlation with SAGAT. This technique requires further studies before to get full acceptance (Jones and Endsley, 2000b).

## **Applicability to C2 Environments**

In contrast to SAGAT techniques, the Reaction Time to On-line probes is a queries technique that does not require the freezing of the situation. From the on-line queries, it captures the operator SA content during the progress of the situation.

This technique requires technological devices to control the occurrence of the probes and to record the time to respond. Consequently, it may be difficult to use the technique in field trials experiments.

The Reaction Time to On-line probes may be best suited to be used within simulated environments where a control over the variables of interest is appropriate. Then, this technique may be used in flight simulators and simulated environments of naval operations. It may also be used in virtual environments simulating infantry operations.

### 3.11 Situation Present Assessment Method (SPAM)

<b>Developers:</b>	Durso, Hackworth, Truitt, Crutchfield, Nikolic, and Manning (1998)
<b>Classification:</b>	Direct / Online
<b>Locus of measurement:</b>	Concurrent

#### Description

SPAM is a real-time probe technique and seems to be the only well described procedure based on real-time probe procedure. Instead of looking for a percentage of correct response as in SAGAT, Durso, Hackworth, Truitt, Crutchfield, Nikolic, and Manning (1998) examine the response time to real-time probes in action scenarios as measures of SA. Two groups of real-time probes were assessing either level 1 or level 2-3 SA. The remaining actions at the end of the scenario were also used as predictors of the final performance. The results show that these variables can predict performance in an air traffic control simulation.

Endsley, Sollenberg and Stein (2000) replicated this experiment without success. In this replication, the response time to real-time probes was not correlated with SAGAT and was slightly correlated with mental workload. The response time variable was not sensitive to the type of display while the SAGAT was. However, the SART and the SMEs performance rating showed a similar lack of sensitivity to this SA-induced condition. SAGAT and SART measures are usually sensitive to fluctuations in SA. Perhaps, the experimental design was not appropriate to test response time sensitivity. However, in wartime versus peacetime scenarios, Jones and Endsley (2000b) found that response time to on-line probes were sensitive and correlated (weak but significant, but not reported by the authors) with SAGAT. Further data are needed to conclude.

#### Psychometric Evaluation

**Sensitivity and Diagnosticity:** Information available for sensitivity but results are rather inconsistent. To our knowledge there was no information available at the time of writing this report for the diagnosticity.

**Selectivity:** To our knowledge there was no information available at the time of writing this report.

**Reliability:** Information available but results are rather inconsistent.

**Validity:** Information available but results are rather inconsistent. Such an instrument needs to be further validated before to get full acceptance.

## **Applicability to C2 Environments**

SPAM is a technique that includes real-time probes to assess the operator SA content. The content is based on the reaction time to the probes included in the scenario. The SPAM does not require the freezing of the experiment and the inclusion of queries during the scenario. In that sense, the realism of the experimental situation is not challenged if the probes are naturally inserted into the situation.

This technique can be used if the experimental setting allows for the inclusion of the probes. However, there is a restriction. As with the Reaction Time to On-line Probes technique, the time to respond to the probe is the dependent variable. Consequently, some data recording devices are required. The SPAM technique may be best suited to be used within simulated environments where control over the variables of interest is appropriate. This is a major point since the SA content is based on the reaction to the presence of a given probes. It is essential to be able to couple the reaction observed with the presence of the probe. This technique may be used in flight simulators and simulated environments of naval operations. It may also be used in virtual environments simulating infantry operations.

### 3.12 Situation Awareness Linked Indicators Adapted to Novel Tasks (SALIENT)

<b>Developers:</b>	Muniz, Stout, Bowers and Salas (1997)
<b>Classification:</b>	Indirect / Offline
<b>Locus of measurement:</b>	Post

#### Description

SALIENT consists of a 27 behavioural self-report checklist. The SA content is inferred from these behaviours, which are claimed to be indicative of good SA. The behaviours are general enough to be used in natural (fields) or technological environments. This approach seems to provide multiple opportunities to evaluate teams based on behaviours associated with team SA.

Retrospective measures depend on two abilities and one attitude: (a) participants' understanding of what is measured, notably situation awareness, (b) the capacity to be self-aware of his own internal thinking both verbal and non-verbal (Taylor and Selcon, 1991), and (c) the will to communicate their real thinking.

Up to date there are a number of validation studies in which some psychometric attributes of SALIENT are tested as a measure of individual SA (e.g., Fink and Major, 2000; see Gawron, 2000) or a measure of team SA (e.g., Muniz, Salas, Stout, and Bowers, 1998). Fink and Major (2000) compared the relative efficacy of SALIENT against two other SA measures, situation awareness probe technique (SAP) and SART. SALIENT showed slightly better psychometric properties and the authors conclude that SALIENT is a very promising measure but that additional research is clearly needed.

#### Psychometric Evaluation

**Sensitivity and Diagnosticity:** To our knowledge there was no information available at the time of writing this report.

**Selectivity:** Correlation with SART is moderate which presumes a certain degree of specificity.

**Reliability:** Very good inter-rater reliability. Originally constructed as a five-factor scale but component analysis reveals a unitary scale.

**Validity:** Good correlation with SART (concurrent) and performance (predictive).

## **Applicability to C2 Environments**

SALIENT is a retrospective measure of SA. Accordingly, it is not used concurrently with the execution of the studied task. As a retrospective measure, it does not alter the execution of the studied task and does not reduce the level of realism of the experimental environments. It can be adapted and used in field trials experiments as well as in simulator environments.

### 3.13 Situation Awareness Rating Scale (SARS)

<b>Developers:</b>	Waag and Houck (1994)
<b>Classification:</b>	Direct-Indirect / Offline
<b>Locus of measurement:</b>	Post

#### Description

SARS involves pilots reporting their assessment of 8 categories that include 31 pre-determined behaviours related to F-15 piloting (Waag, and Houck, 1994). In this way, SARS can be viewed as a comparative measure based on the inferences made from these behaviors. This scale is closely bound to the F-15 aircraft piloting and it can hardly be generalized to other tasks. It was nevertheless innovative since pilots were required to self-report their own SA but also to evaluate their colleagues SA.

Operator- or self-rating is a simple, low cost technique in which operators rate their SA, usually on some form of scale. Several concerns arise in using this method, however, including that operators rarely know how incomplete their perception of a situation is, that performance on the task (e.g., simulation of air combat) will drive perception of their SA, and that over-generalization is a common side-effect resulting in a lack of detailed feedback (Endsley, 1988a). Endsley (1995a) further argues that self-rated SA is more a measure of operator confidence, and less a measure of true awareness.

Waag and Houck (1994) evaluated the psychometric properties of SARS from three perspectives: self, peers and supervisor ratings. 205 F15 pilots took part in this validation study. SARS was shown to possess very good psychometric properties

#### Psychometric Evaluation

**Sensitivity and Diagnosticity:** To our knowledge there was no information available at the time of writing this report.

**Selectivity:** Specific to F15 piloting but correlated with mission performance.

**Reliability:** Internal consistency and inter-rater reliability are very good.

**Validity:** Correlations between rating perspectives are moderate to high. There is no data yet about the correlation of SARS with other measures.

#### Applicability to C2 Environments

SARS is a retrospective measure of SA and is therefore not used concurrently with the execution of the studied task. Consequently, it does not alter the execution of the studied task and does not reduce the level of realism of the experimental environments.

Rather, it is used to measure the pilot's SA content. However, the framework under the measurement technique can be applied to the other C2 environments (see PSAQ for infantry environments). It can be adapted and used in field trials experiments as well as in simulator environments.

### **3.14 Participant Situation Awareness Questionnaire (PSAQ)**

<b>Developers:</b>	Matthews, Pleban, Endsley, and Strater (2000)
<b>Classification:</b>	Direct / Offline
<b>Locus of measurement:</b>	Post

#### **Description**

Most existing subjective measures of SA were developed for aviation and therefore are unsuitable for infantry. PSAQ was developed to fill that need. The PSAQ is administered at the conclusion of an exercise, either in the field or simulation. Participants have to rate on a five-point scale their awareness of several aspects of a situation. They must answer question such as “Please circle the number that best describes how aware of the evolving situation you were during the scenario?” This measure is still in development. This instrument should be used in conjunction with the SABARS and the SAGAT-MOUT.

#### **Psychometric Evaluation**

To our knowledge there was no information available at the time of writing this report.

#### **Applicability to C2 Environments**

PSAQ is a retrospective measure of SA and is therefore not used concurrently with the execution of the studied task. Consequently, it does not alter the execution of the studied task and does not reduce the level of realism of the experimental environments. Actually, it is used to measure the soldier’s SA content in infantry operations. However, the framework under the measurement technique can be applied to the other C2 environments (see SARS for pilots). It can be adapted and used in field trials experiments as well as in simulator environments.

This page intentionally left blank.

### 3.15 Situation Awareness Rating Technique (SART)

<b>Developers:</b>	Taylor (1990)
<b>Classification:</b>	Direct / Offline
<b>Locus of measurement:</b>	Post

#### Description

SART is a subjective measurement technique that concentrates on assessing an operator's knowledge in three areas: (1) demands on attentional resources, (2) supply of attentional resources, and (3) understanding of the situation. Thus, it considers operators' perceived workload as defined by attention capacity (supply minus demand) in addition to their perceived understanding (awareness) of the situation in question.

SART provides a validated and practical subjective rating tool for the measurement of SA, based on personal construct dimensions associated with SA (Taylor, 1990). SART gives ratings of subjective dimensions associated with SA, based on 10 aircrew personal constructs, representing the psychological dimensions of SA. The constructs dimensions were elicited from aircrew, utilising the Repertory Grid technique, through consideration of aircrew descriptions of tactical flight situations involving SA.

The structure of the construct dimensions has been interpreted as comprising three related conceptual groups, which form the principal dimensions of SART, namely: (i) Demand for Attentional Resources or D (complexity, variability, instability); (ii) Supply of Attentional Resources or S (arousal, concentration, division of attention, spare mental capacity); and (iii) Understanding of the situation or U (information quality, information quantity, familiarity). SART provides 7-point Likert rating scales for the three principal derived dimensions (3-D SART), and for the 10 sub-ordinate, original personal construct dimensions (10-D SART). There is also new anchored version of the 3-D SART, with antonyms and synonyms at the scale poles for easier interpretation, based on the 10-D sub-ordinate dimensions.

SART is, with SAGAT, the most common measure of SA. SART is a subjective measure and, as such, suffers from the inherent problems of all subjective measures. However, SART is very easy to administer and does not require costly and time consuming development of queries or implementation. SART can be used in all sorts of situations and has been extensively validated (as opposed to most measures).

#### Psychometric Evaluation

**Sensitivity and Diagnosticity:** SART is sensitive to differences in performance of aircraft attitude recovery tasks and learning comprehension tasks (Selcon and Taylor, 1991). SART is also sensitive to pilot experience (Selcon, Taylor, and Koritsas, 1991). One weakness with self-ratings generally, is their inability to take account of what is not known by the ratee, known as the "6 o'clock problem". SART appears to offer

diagnosticity in that it is significantly related to performance measures of rule and knowledge-based decision-making (Selcon and Taylor, 1990).

**Selectivity:** However in terms of specificity it is highly correlated to workload measures. Pew (1995) criticises this technique, stating that, at worst, it produces a measure of operator confidence or preferences, and at best, it produces what the researcher is interested in obtaining anyway, as a guide to further study.

**Reliability:** To our knowledge there was no information available at the time of writing this report.

**Validity:** Good. SART has been shown to have some useful predictive power, accounting for some 30-40% of the variance in performance data in a simulated ATC task (Taylor, Selcon, and Swinden, 1995).

### **Applicability to C2 Environments**

SART is a retrospective measure of SA and is therefore not used concurrently with the execution of the studied task. Consequently, it does not alter the execution of the studied task and does not reduce the level of realism of the experimental environments. It can be adapted and used in field trials experiments as well as in simulator environments.

### 3.16 Situation Awareness Subjective Workload Dominance Technique (SA-SWORD)

<b>Developers:</b>	Vidulich (1989)
<b>Classification:</b>	Direct / Offline
<b>Locus of measurement:</b>	Post

#### Description

SA-SWORD uses the same paradigm used with the SWORD technique, but with different instructions, in order to assess situation awareness (Vidulich and Hughes, 1991). SA-SWORD allows operators to make pairwise comparative ratings of competing design concepts along a continuum that expresses the degree to which one concept entails less workload/SA than the other. SA-SWORD has participants provide a comparative preference for displays on a 9-point scale, based on their beliefs about the amount of SA provided by each. One problem with this technique is that it is difficult to assert conclusively whether subjective preference or higher SWORD ratings of SA are being indicated by participants, given that the 'preferred' display will almost invariably rate the higher SA (Endsley, 1995a).

#### Psychometric Evaluation

**Sensitivity and Diagnosticity:** SWORD is a paradigm developed to measure mental workload and has a high sensitivity (e.g., sensitive to differences in tracking tasks and color displays). To our knowledge there was no information available for diagnosticity at the time of writing this report.

**Selectivity:** It is a measure of workload related to SA but it does not provide information about the SA content.

**Reliability:** SWORD has a good test-retest reliability (e.g., +.937) and good interrater reliability in comparing a design/display (Endsley, 1995a).

**Validity:** To our knowledge there was no information available at the time of writing this report.

#### Applicability to C2 Environments

SA-SWORD is a retrospective measure. However, it does not measure the SA content in a given situation. The goal with this measure is to compare different possible displays designed to support SA. The comparisons are made according to the level of SA and workload related to the different displays. The comparison is made after the use of different displays. Consequently, it does not alter the execution of the studied task and does not reduce the level of realism of the experimental environments. It may

be appropriate in field trials and simulator environments. However, the goal of this measure is to compare potential displays. Accordingly, naval and piloting environments may be more appropriate.

### 3.17 Map Reconstruction

<b>Developers:</b>	-----
<b>Classification:</b>	Indirect / Offline
<b>Locus of measurement:</b>	Post

#### **Description**

A recall situation technique simply asks a person to rebuild the map of a situation at a given time. For example, an officer may be asked to indicate the position of his infantrymen on a situation map ten minutes before a significant event. This technique can be used as a prospective measure. However, the map constructed can be compared with a map representing the exact situation. There is no evidence that this technique measures SA. However, the extracted information is what is considered as part of SA.

#### **Psychometric Evaluation**

To our knowledge there was no information available at the time of writing this report.

#### **Applicability to C2 Environments**

The Map Reconstruction technique is a retrospective measure of SA and is therefore not used concurrently with the execution of the studied task. Consequently, it does not alter the execution of the studied task and does not reduce the level of realism of the experimental environments. It can be adapted and used in field trials experiments as well as in simulator environments. Actually, it may be more appropriate in infantry operations. However, it may be possible to reconstruct displays and screens representations as well.

This page intentionally left blank.

### **3.18 China Lake Situation Awareness (CLSA)**

<b>Developers:</b>	Adams (1998)
<b>Classification:</b>	Direct / Online
<b>Locus of measurement:</b>	Concurrent

#### **Description**

The China Lake Situational Awareness (CLSA) is a five-point rating scale. CLSA, adapted from Endsley's work on SAGAT, is administered like the Bedford Workload Scale. Contrarily to SAGAT, the Bedford Workload Scale had been developed and field-tested. CLSA was designed at the Naval Air Warfare Center at China Lake to measure SA in flight (Adams, 1998). An important concern with this technique is that data collection does not compromise safety. Validation and development of CLSA is still in work.

#### **Psychometric Evaluation**

The psychometric evaluation is in progress.

#### **Applicability to C2 Environments**

The CLSA technique is an on-line rating method. Therefore, it is essential that the situation allow the inclusion of this rating task into the execution of the studied SA task. The execution of the CLSA technique may alter the execution of the studied task and may reduce the level of realism of the experimental environments.

This page intentionally left blank.

### 3.19 IMPLIcit Situation Awareness toolkit (IMPSA)

<b>Developers:</b>	Croft and Banbury (QinetiQ; 2000)
<b>Classification:</b>	Indirect / Offline
<b>Locus of measurement:</b>	Post

#### Description

Recently, researchers have become increasingly interested in developing tests to measure knowledge that is less available to conscious introspection or that is completely unconscious. The development of these implicit tests/probes underscores the possibility that existing methodologies for measuring SA may not be tapping all of a system operator's available knowledge. Current SA measures may be focusing too heavily on what an operator can access from conscious memory, and can articulate. There are some advantages of implicit memory probes: (i) Longer retention interval – Testing can be performed post hoc, rather than during a trial; (ii) Independent of locus of attention – Testing of poorly attended objects can be performed; and (iii) Incidental test – Testing can be performed covertly, and therefore is less prone to bias. By definition, then, implicit knowledge tests provide us with a means of accessing information that is not available using conventional explicit memory tests. All implicit tests are 'incidental', in that participants are not aware that their memory or knowledge is being probed.

Most of the research that has been conducted into implicit memory and learning has been carried out in academia and has had limited practical application. The current version of the IMPSA toolkit is a collection of the basic laboratory methods that demonstrated applied potential, coupled with two modified tests that were developed during the course of this research. The toolkit will enable users to design a test for measuring implicit SA by taking them through the design and construction stages in a self-explanatory manner. When users first enter the toolkit, they will be provided with some background information relating to the development of the toolkit, information as to funding sources, and a list of credits. On entering the toolkit proper, the user will be provided with detailed information relating to the topics of SA, situation knowledge, and implicit/explicit memory and knowledge.

#### Psychometric Evaluation

The psychometric evaluation is in progress.

#### Applicability to C2 Environments

The IMPSA toolkit is used after the completion of the studied SA task. Consequently, it does not alter the execution of the studied task and does not reduce the level of

realism of the experimental environments. It can be adapted and used in field trials experiments as well as in simulator environments.

## 3.20 Situational Awareness Supervisory Rating Form (SASRF)

<b>Developers:</b>	Caretta, Perry and Ree (1996)
<b>Classification:</b>	Indirect / Online
<b>Locus of measurement:</b>	Concurrent

### Description

SASRF is a subjective rating scale specific to F-15 tactical missions. The rating form has 31 items that range from general traits to tactical employment. Supervisors and peers make the ratings.

### Psychometric Evaluation

**Sensitivity and Diagnosticity:** To our knowledge there was no information available at the time of writing this report.

**Selectivity:** SASFR is specific to F-15 tactic missions.

**Reliability:** To our knowledge there was no information available at the time of writing this report.

**Validity:** To our knowledge there was no information available at the time of writing this report.

### Applicability to C2 Environments

The SASRF technique is a rating method used by supervisors and peers during the execution of the task. The rating technique involves the inclusion of an observer into the task environments. Consequently, the limitation of this technique is related to the possibility of including someone in the environment without altering it. This technique can be used if the inclusion of the observer does not reduce the level of realism of the experimental environments.

The SASRF technique is specific to the F-15 piloting. There are other rating techniques available such as SABARS for infantry operations. The observer rating is appropriate in simulation environments. A benefit of this technique is that the operator's task is not altered. Consequently, if the presence of the observer is possible and does not reduce the level of realism of the simulation, the observer rating technique can be adapted to most of the simulation environments.

This page intentionally left blank.

## 4. Evaluation of the Team SA Measurement Tools

---

The following section describes a number of measures used to evaluate the SA of teams; in terms of either the SA of the team as a whole, or the SA of the individuals that comprise the team. Once again, in addition to a general description of the measure, information pertaining to its sensitivity, diagnosticity, selectivity, validity, reliability and applicability to C2 environments is provided. However, given the lack of development of these approaches, very little of this information is available for most of these factors.

### 4.1 DUO WOMBAT Crew Resource Management Test

**Developers:** LaRoche and Roscoe (Roscoe, Corl and LaRoche, 2001)

**Classification:** Indirect / Online

**Locus of measurement:** Not Applicable

#### Description

WOMBAT is a modern psychological assessment tool for selecting complex system operators such as pilots, air traffic controllers, ship operators, nuclear-plant operators, in fact anyone in charge of complex operations involving multiple concurrent inputs and response alternatives. WOMBAT may have the look and feel of a video game. However, it is much more than a complex game; it is a powerful, culture-independent PC-based set of tasks. These include target tracking, spatial orientation, pattern recognition and working memory. It is claimed by the developers to be the ideal system for measuring situational awareness under stress. The procedure is simple. The applicant sits in front of a desktop computer and interacts with it using a special WOMBAT console. The console has two standard joysticks and a custom keypad. Performance measures have the advantage of being objective, usually non-invasive, and easy to collect.

To measure how well crew resources are managed, the WOMBAT-CS was expanded into the DuoWOMBAT Crew Resource Management Test. Duo WOMBAT measures the aptitude to operate a complex system (as solo WOMBAT) and also the team aspects of performance. It has been developed to measure and enhance team performance through effective team resource management.

#### Psychometric Evaluation

To our knowledge there was no information available at the time of writing this report.

## **Applicability to C2 Environments**

Duo WOMBAT is a test battery with its own experimental setting including displays, joysticks and consoles. The measurement of the SA concept with Duo WOMBAT is an experiment itself with its own devices, protocols and settings. It can only identify, from its own experimental setting, how well individuals work together to form good SA.

Because of the nature of Duo WOMBAT, it cannot be inserted into infantry, air and navy operations during field trials. The activities required by Duo WOMBAT may alter considerably the realism of the trials. Also, the inclusion of Duo WOMBAT test battery may not be appropriate within simulation environments. Consequently, it cannot be coupled with experimental settings.

Duo WOMBAT can be used before field trials or experiments with simulators to determine a priori the human potential to develop a good level of SA when in two-person teams.

## 4.2 Crew Awareness Rating Scale (CARS)

<b>Developers:</b>	McGuinness and Foy (1999, 2000)
<b>Classification:</b>	Direct / Offline
<b>Locus of measurement:</b>	Post or concurrent

### Description

The Crew Awareness Rating Scale (CARS) is a generic 8-part questionnaire addressing both SA content and processing with regards to four separate functions (perception, comprehension, projection, integration). Eight questions investigate these four contents and these four processes categories. Each item is answered according to a subjective 4-point scale: (a) definitely +, (b) probably +, (c) probably -, and (d) definitely -. This allows the experimenter to distribute the answers into a 2 by 2 grid, which allows data to be collected for both verifiable perceived SA (correct or incorrect) and personal certitude about this SA (certain or uncertain) with regards to the contents and processes of the four levels.

CARS provides subjective SA ratings and is administered to individuals. It is sometimes referred to as a team SA measure (Breton and Rousseau, 2001) due to its frequent use in infantry and crew situations but CARS does not evaluate the team aspects of SA.

### Psychometric Evaluation

**Sensitivity:** CARS is sensitive to the implementation of a traffic alert and collision avoidance system (TCAS), revealing an increased degree of SA (Foy and McGuinness, 2000). However, even though the use of battlespace digitization improved objective SA, subjective ratings measured with CARS were not different whether conventional or digital mapping of the battlespace was employed (McGuinness and Ebbage, 2001).

To our knowledge there was no information available at the time of writing this report for the other psychometric properties.

### Applicability to C2 Environments

The CARS technique is an on-line rating method. As such, it is essential that the situation allow for the inclusion of this rating task into the execution of the studied SA task. The execution of the CARS technique may alter the execution of the studied task and may reduce the level of realism of the experimental environments. The basic format of CARS is generic enough to adapt the tool to field trials and simulation environments. It is actually restricted to air transport crew. But it could be adapted in used in infantry and naval environments involving crews.

This page intentionally left blank.

### 4.3 Crew Situational Awareness (CSA)

**Developers:** Mosier and Chidester (1991)

**Classification:** Direct / Online

**Locus of measurement:** Concurrent

#### Description

Expert observers rate crew co-ordination and performance errors and then develop information transfer matrices identifying time and source of item requests (prompts) and verbalised responses. This method requires frequent communication among aircrew members. Further, it requires a team of experts to develop the information transfer matrices.

#### Psychometric Evaluation

*Sensitivity:* CSA is sensitive to error types and decision prompts.

To our knowledge there was no information available at the time of writing this report for the other psychometric properties.

#### Applicability to C2 Environments

The CSA technique appunderscores the level of SA crews are exhibiting through the study of their communication and performance errors. As such, this technique requires frequent communication between team members. One problem might be that team members might be compelled to verbalise the logic of their behaviours and actions. For this reason, this technique may interfere with the non-verbal aspects of the on-going processes.

The CSA technique may be appropriate for field trials testing. However, since the operator might feel compelled to justify his behaviours concurrently with the execution of the actions, there is a high probability that the execution of task may be altered. In that sense, the use of this technique in field trials may reduce considerably the level of realism required by these experiments. The alteration of the studied task seems to be the only major cost related to this technique. It can be used in simulation environments.

This page intentionally left blank.

## 4.4 Integrative Methods Approach

**Developers:** Hauland (2002; see also Andersen and Hauland, 2000)

**Classification:** Direct / Online

**Locus of measurement:** Concurrent

### Description

IMA is process-oriented and combines the analysis of eye movement (visual attention) and a variety of behavioural data such as verbal communication to pre-defined probe events embedded in the simulated tasks. Probe events were designed to require visual attention and coordination within and between team members. Visual acquisition was measured with the dwell time to pre-defined areas of interest.

This approach uses a test battery that integrates data from typical individual SA measures to coordination (co-occurrence) of actions and communication between individuals. The variables and measures included in IMA may vary across situations. IMA has been tested with simulated tasks in ATC and nuclear power plant environments.

### Psychometric Evaluation

To our knowledge there was no information available at the time of writing this report on the psychometric properties of IMA as a battery of SA-related measures.

### Applicability to C2 Environments

This technique relies on, in part, EPOG technologies (described in the previous section, p. 28). However, because of the reliance on the technological devices to collect the data, this type of measure is not applicable to field trials. Most of eye-trackers ask for a certain level of stability of the operator's head as well as for the technological settings. They can also considerably limit the operator's mobility. Instead, eye-trackers are better suited to be coupled with simulators. Ideally, eye-trackers should be used in experimental settings including displays, consoles, radar screens, etc. The operator should be seated in front of those devices to allow a certain level of stability required by the eye-tracker equipments.

An additional limitation of this technique relates to the study of verbal communication between crew members insofar as team members might feel compelled to verbalise the logic of all their behaviours and actions. For this reason, this technique may interfere with the non-verbal aspects of the on-going processes.

This page intentionally left blank.

## 4.5 Targeted Acceptable Responses to Generated Events or Tasks (TARGETs)

**Developers:** Fowlkes, Lane, Salas, Franz and Oser (1994)

**Classification:** Direct / Online

**Locus of measurement:** Concurrent

### Description

Fowlkes et al. (1994) developed the TARGETs methodology to assess team behaviour in CRM contexts (see also Line/LoS checklist; Helmreich and Foushee, 1993). Expected responses are scripted *a priori* for each scenario event and team responses are evaluated on the basis of their match with the scripted 'targeted acceptable responses'. However, the TARGETs methodology provides only a frequency count of the team behaviours and does not give any indication of the underlying team processes.

### Psychometric Evaluation

To our knowledge there was no information available at the time of writing this report.

### Applicability to C2 Environments

The TARGETs approach seeks to appunderscore the level of SA crews are exhibiting through the study of their communication and performance errors. As such, this technique requires frequent communication between team members. One problem might be that team members might be compelled to verbalise the logic of their behaviours and actions. For this reason, this technique may interfere with the non-verbal aspects of the on-going processes.

The TARGETs approach may be appropriate for field trials testing. However, since the operator might feel compelled to justify his behaviours concurrently with the execution of the actions, there is a high probability that the execution of task may be altered. In that sense, the use of this technique in field trials may reduce considerably the level of realism required by these experiments. The alteration of the studied task seems to be the only major limitation of this technique. It can be used in simulation environments.

This page intentionally left blank.

## 4.6 Situation Awareness and Workload In-flight Measure (SAW-IM)

**Developers:** Banbury, Hoermann, Soll and Dudfield (in review)

**Classification:** Indirect / Offline

**Locus of measurement:** Post

### Description

The SAW-IM measure is derived from the China Lake Situation Awareness (see CLSA in the previous section) measure. The CLSA involves test pilots making SA ratings on a five-point scale during relatively benign portions of the test-flight. Each point on the five-point scale refers to a relatively complex description of the SA and workload experienced. Although appropriate for experienced test-pilots who presumably would be well-trained using the scale, it was thought that this approach was overly complex for non-test pilot aircrew. Furthermore, it was decided to create a separate scale for workload and SA, rather than combine them, as it was thought that confounding both constructs on the same scale was an important limitation of the CLSA approach.

In order to simplify the rating procedure (and therefore reduce the obtrusiveness of the rating procedure), a 3-point scale was adopted for the SA and workload ratings. For the SA scale, the items were “behind the aircraft”, “with the aircraft”, and “ahead of the aircraft”. For the workload scale the items were “underload”, “optimal”, and “overload”. At various times throughout the simulator session, participants are prompted by the experimenter to rate their own awareness level and their own workload level.

In addition to the self-ratings, participants are asked to rate their perception of the other crewmember’s awareness level and their perception of the other crewmember’s workload level (i.e. peer-ratings). This is a novel aspect of the SAW-IM insofar as other measures of workload and SA have focused on self-ratings only. The rationale for the peer-ratings is to probe participants’ ability to recognize losses (or gains) in SA, or increases (or decreases) in workload of the other crewmember. Such recognition of these attributes is thought to be important for effective TSA. In order to reduce the obtrusiveness of the measure, ratings are elicited at relatively benign segments of the flight, and participants are asked to rate SA and workload for the period of time since their last rating (or from the start).

### Psychometric Evaluation

The psychometric evaluation is in progress.

## **Applicability to C2 Environments**

The SAW-IM technique is an on-line rating method. As such, it is essential that the situation allows the inclusion of this rating task into the execution of the studied SA task. The execution of the SAW-IM ratings may alter the execution of the studied task and may reduce the level of realism of the experimental environments.

## 4.7 Factors Affecting Situation Awareness (FASA)

**Developers:** Banbury, Hoermann, Soll and Dudfield (in review).

**Classification:** Indirect / Offline

**Locus of measurement:** Post

### Description

There are a number of factors that have been argued to influence the process of acquiring and maintaining SA (Endsley, 1995b). First, individuals may vary in their ability to acquire SA as a function of their cognitive abilities, which in turn may be influenced by innate abilities, experience and training. In addition, individuals may possess certain preconceptions and objectives that can influence their perception and interpretation of their environment. However, most of the existing measures of SA focus solely on measuring the level of SA in terms of ‘product’ (i.e. participants’ awareness of key SA elements at one moment in time). Critically, they do not focus on the measurement of SA in terms of ‘processes’ (i.e. the processes involved in situation assessment undertaken to produce a representation in memory, or product). Given that the present study was concerned with improving the processes that contribute to SA through training, a more diagnostic measure (i.e. a measure that is concerned with the cognitive ‘processes’, rather than the resultant ‘product’) was required to assess the efficacy of this training.

An exhaustive search of the cognitive factors thought to be important to the acquisition and maintenance of SA was undertaken using a large number of laboratory and field-based sources (see ESSAI, 2001). Several of the cognitive factors were discounted because it was questionable whether respondents would have any insight into these psychological processes. From the list of remaining factors, a number of questions were generated to elicit a response from pilots as to how important they feel these factors to be. These were then incorporated into the FASA questionnaire. In addition, a scoring regimen for this scale was created so that each respondent would have a score to reflect how susceptible they are to factors that are thought to influence the acquisition and maintenance of sufficient SA.

The FASA questionnaire is divided into five sub-scales, one of which concerns Inter-Personal Dynamics. Non-verbal communication is an important information channel between team members (reference). Non-verbal cues can be divided into; physical characteristics; kinesic cues (i.e. body language); paralanguage (e.g. intonation, expression); and proxemic cues (i.e. proximity). Clearly, understanding the impact of these cues should allow individual’s to make better judgements of their team-mate’s level of workload or SA. In light of these findings, questions in this sub-scale relate to participants’ knowledge of non-verbal communication.

The scale is constructed in a five point Likert format: Strongly Disagree, Disagree, Neutral, Agree, and Strongly Agree. There are six questions per sub-scale making a

total of 30 questions. The scoring of the scale is such that the higher the score, the less the participant is susceptible to factors that negatively affect acquisition and maintenance of SA.

Ideally, the FASA questionnaire is presented twice immediately after the experimental session. For the first presentation, participants are instructed to indicate their agreement or disagreement with the statements (i.e. a measure of their actual attitudes during the simulator session). However, for the second presentation participants are instructed to respond to the statements in terms of how they would have liked to behave ideally. The rationale for this manipulation is that the effects of training may not become immediately apparent. In other words, the effects of the training may not be manifest in observable changes in skill or knowledge. By asking participants to rate their ideal behaviour, it may be possible to at least observe any changes in attitude.

### **Psychometric Evaluation**

The psychometric evaluation is in progress.

### **Applicability to C2 Environments**

FASA is a retrospective measure of SA, and as such is not used concurrently with the execution of the studied task. Consequently, it does not alter the execution of the studied task and does not reduce the level of realism of the experimental environments. It can be adapted and used in field trials experiments as well as in simulator environments. Most importantly, it is one of the few measures to examine the impact of the cognitive processes underlying SA (i.e. it can be used as a diagnostic tool to identify the causes of low or high levels of SA).

## 4.8 Information Propagation in Team Environments (IPTE)

**Developers:** Banbury and Howes (2001)

**Classification:** Direct / Online

**Locus of measurement:** Concurrent

### Description

Banbury and Howes (2001) proposed an objective measure of TSA. Consistent with other definitions of team SA, they conceive overall TSA, whereby each team member has a specific set of elements about which he is concerned. Moreover, an individual's set of SA elements is, in part, determined by each member's responsibilities within the team. Clearly, some overlap between each team member's SA requirements will be present, and it is this subset of information that constitutes much of team co-ordination and communication. As such, the quality of team members' SA of shared elements (i.e. as a state of knowledge) can serve as an index of team co-ordination or effectiveness.

There are a number of ways that TSA could be assessed. One is that questionnaires could be designed to expose (a) types of knowledge held by individuals; and (b) what they know about other's goals and knowledge. Information collected in category (a) could then be used to assess the general accuracy of information collected in category (b). The accuracy of the evaluation could be increased with iterations in which individuals are asked about the accuracy of what other's think they know about their goals and knowledge.

A second, more novel, approach to assessing team SA would be the deliberate propagation of information tokens, or 'seeds', in the team environment and the subsequent observation of their trajectory within that environment. Clearly, the information seeds must be both salient enough to allow observation, and critical enough to observe the effect on an individual team members' decision making. In, other words, a particular outcome of a particular team member's decision-making would indicate that the information has reached the intended recipient.

Using this approach, the efficacy of team processes (e.g. communication, monitoring and so on) could be evaluated. For example, the following aspects of information flow can be assessed: where it goes, who it reaches, how long it takes to get there and whether the information content is maintained correctly.

Although these TSA methodologies are at present speculative, the means through which it can be evaluated have already been widely researched: system logging, interviews, questionnaires of expert observation can all ascertain the use of the information 'seed'. Importantly the evaluation could be conducted in naturalistic settings. The authors are currently developing this technique further.

## **Psychometric Evaluation**

The psychometric evaluation is in progress.

## **Applicability to C2 Environments**

This technique uses responses to embedded probes to appunderscore the level of SA a team is exhibiting. As a result, the technique does not require the freezing (i.e. interruption) of the situation. However, this technique requires technological devices to control the occurrence of the probes and to record the time to respond, and as such, makes it difficult to use the technique in field trials experiments.

The IPTE technique may be best suited to simulated environments where control over the variables of interest is possible. In this form, this technique may be used in flight simulators and simulated environments of naval operations. It may also be used in virtual environments simulating infantry operations.

## 5. The Distribution of the SA Measurement Tools among C2 Environments

---

Table 3 shows the distribution of the SA measurement tools among the different C2 environments. The tools are assigned to a specific environment according to their actual use (even if some can be adapted to be used in other settings). Some tools (in bold) can be used in all situations (field-simulation and air-navy-infantry). The use of others is kept specific (in italic) to a given environment.

Some observations can be made about the distribution of the SA measurement tools among the C2 environments shown in Table 3. First, most of tools are used in simulation environments in comparison with field trials. Field trials are time consuming, very costly and the control over the experiment is somehow difficult. One of the main reasons to perform field trials is because the real-life environment provides a better venue to validate technological systems, new training programs, etc. However, developments in technology have facilitated the creation of powerful, high fidelity simulators that can provide high levels of realism without compromising the control a researcher has over the experiment. Thus, because new simulator technologies are providing environments with greater realism and more control than in the past, it is not surprising that most of the SA tools are developed more for simulated than real environments.

Second, it is interesting to see that the tools are almost evenly distributed among the air, navy and infantry environments. Two reasons may explain this type of distribution. A) SA is evenly important in all these environments and B); these environments have constraints that force the use of specific SA measurement tools. For instance, in air environments, it may be hard to include the presence of an observer in F-15 cockpit or to use a mouse (Force to Ask) to perform the piloting task. In infantry environments, the battlefield dimension can restrict, for instance, the data recording process. The constraints related to the C2 environments make necessary the development of various SA measurements to fit to the different environment constraints.

A third is that few tools are general enough to be used in all C2 environments in field trials and simulation environments. SA is largely dependent of the situation in which the SA task is performed. Since every situation presents its own set of constraints, it is not surprising to observe that few tools are general enough to fit in all settings. These all-purpose tools are the WOMBAT test battery, the Think Aloud technique and three retrospective measures, SALIANT, SART and IMPSA. WOMBAT is not seen as a typical SA measurement tool. WOMBAT is a test battery to evaluate the individual's potential to develop good SA content. Among these all-purpose tools, only the Think Aloud technique is used concurrently with the execution of the task. It is a very simple technique that does not ask for sophisticated data recording process. SALIANT, SART and IMPSA can also be used in all the different environments since they are retrospective measures (used after the execution of the task) that can easily be adapted to different situations.

**Table 3. Distribution of the SA measurement tools among the C2 environments.**

C2 Situations	Field Trials		Simulation	
	Individual	Team	Individual	Team
Air	<ul style="list-style-type: none"> <li>• <b>Wombat</b></li> <li>• <b>Think Aloud</b></li> <li>• <b>SALIENT</b></li> <li>• <i>SARS</i></li> <li>• <b>SART</b></li> <li>• SA-SWORD</li> <li>• <i>CLSA</i></li> <li>• <b>IMPSA</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>CSA</b></li> <li>• <b>TARGETs</b></li> <li>• <b>SAW-IM</b></li> <li>• <b>FASA</b></li> <li>• <i>CARS</i></li> <li>• <b>Duo Wombat</b></li> </ul>	<ul style="list-style-type: none"> <li>• Eye Trackers</li> <li>• EEG</li> <li>• <b>Wombat</b></li> <li>• <b>Think Aloud</b></li> <li>• SAGAT</li> <li>• <i>SAGAT Partial Blanking</i></li> <li>• RT to on-line probes</li> <li>• SPAM</li> <li>• <i>CLSA</i></li> <li>• <b>IMPSA</b></li> <li>• <i>SASRF</i></li> <li>• <b>SALIENT</b></li> <li>• <i>SARS</i></li> <li>• <b>SART</b></li> <li>• SA-SWORD</li> </ul>	<ul style="list-style-type: none"> <li>• <i>CARS</i></li> <li>• <b>CSA</b></li> <li>• <b>SAW-IM</b></li> <li>• <b>FASA</b></li> <li>• IPTE</li> <li>• Integrated Methods</li> <li>• <b>Duo Wombat</b></li> </ul>
Naval	<ul style="list-style-type: none"> <li>• <b>Wombat</b></li> <li>• <b>Think Aloud</b></li> <li>• <b>SALIENT</b></li> <li>• <b>SART</b></li> <li>• SA-SWORD</li> <li>• <b>IMPSA</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Duo Wombat</b></li> <li>• <b>CSA</b></li> <li>• <b>TARGETs</b></li> <li>• <b>SAW-IM</b></li> <li>• <b>FASA</b></li> </ul>	<ul style="list-style-type: none"> <li>• Eye Trackers</li> <li>• EEG</li> <li>• <b>Wombat</b></li> <li>• <i>Force to Ask</i></li> <li>• <b>Think Aloud</b></li> <li>• SAGAT</li> <li>• RT to on-line probes</li> <li>• SPAM</li> <li>• <b>SALIENT</b></li> <li>• <b>SART</b></li> <li>• SA-SWORD</li> <li>• <b>IMPSA</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>CSA</b></li> <li>• <b>SAW-IM</b></li> <li>• <b>FASA</b></li> <li>• IPTE</li> <li>• Integrated Methods</li> <li>• <b>Duo Wombat</b></li> </ul>
Infantry	<ul style="list-style-type: none"> <li>• <b>Wombat</b></li> <li>• <b>Think Aloud</b></li> <li>• <i>SABARS</i></li> <li>• <b>SALIENT</b></li> <li>• <i>PSAQ</i></li> <li>• <b>SART</b></li> <li>• <i>MAP Reconstruction</i></li> <li>• <b>IMPSA</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>CSA</b></li> <li>• <b>TARGETs</b></li> <li>• <b>SAW-IM</b></li> <li>• <b>FASA</b></li> <li>• <b>Duo Wombat</b></li> </ul>	<ul style="list-style-type: none"> <li>• Eye Trackers</li> <li>• EEG</li> <li>• <b>Wombat</b></li> <li>• <i>SABARS</i></li> <li>• <b>Think Aloud</b></li> <li>• SAGAT</li> <li>• <i>SAGAT-MOUT</i></li> <li>• RT to on-line probes</li> <li>• SPAM</li> <li>• <b>SALIENT</b></li> <li>• <i>PSAQ</i></li> <li>• <b>SART</b></li> <li>• <i>MAP Reconstruction</i></li> <li>• <b>IMPSA</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>CSA</b></li> <li>• <b>SAW-IM</b></li> <li>• <b>FASA</b></li> <li>• IPTE</li> <li>• Integrated Methods</li> <li>• <b>Duo Wombat</b></li> </ul>

Finally, the fourth observation is that most of tools that are specific to an environment (in italic in Table 3) are related to infantry and flight situations (see Table 4). The particularities (battlefield, tempo) of the infantry operations and air environments make necessary the development of such specific SA tools.

**Table 4. Distribution of the specific SA measurement tools among the C2 environments.**

C2 Situations	Field Trials	Simulation
Air	<ul style="list-style-type: none"> <li>• <i>SARS</i></li> <li>• <i>CARS (T)</i></li> <li>• <i>CLSA</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>SAGAT Partial Blanking</i></li> <li>• <i>SARS</i></li> <li>• <i>CARS (T)</i></li> <li>• <i>CLSA</i></li> <li>• <i>SASRF</i></li> </ul>
Naval	<ul style="list-style-type: none"> <li>• <i>SART</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>Force to Ask</i></li> </ul>
Infantry	<ul style="list-style-type: none"> <li>• <i>SABARS</i></li> <li>• <i>PSAQ</i></li> <li>• <i>MAP Reconstruction</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>SABARS</i></li> <li>• <i>SAGAT-MOUT</i></li> <li>• <i>PSAQ</i></li> <li>• <i>MAP Reconstruction</i></li> </ul>

Breton and Rousseau (2001) have listed, in a literature review, 26 different definitions of SA. Obviously, the problem of measuring a concept must be tightly coupled with the problem of defining the concept. To adequately measure a concept, one must know exactly what the concept is. Then, considering the large number of SA definitions found in the literature, it is not surprising to find as much as SA measurement tools in the literature. Another reason explaining this large number of SA measures is that different types of measures are required to measure SA in different situations.

The impressive number of SA measurement tools underscores the problem of selecting the appropriate measure for a given situation. According to Breton and Rousseau, a fruitful strategy to select a measure for SA is to define this concept according to the situation in which the task is performed and then chooses a tool based on the characteristics of the situation. They propose a three-step model to select a SA measurement tool according to the objective of the measure, and the criteria derived from the situation in which the measure is taken, the content identified from verbal reports or inferred from observation and the person or group of persons performing the assessment of the situation.

This page intentionally left blank.

## **6. Applicability of Individual Measures of SA to TSA**

---

Situation awareness can be defined as an internal state of mind on the situation in which we are involved. We see the world around us in individual terms, based on our cultural background, education, and experiences, not to mention the strengths and limitations of our senses. Situation awareness can be characterized as a dynamic mental model of our operating environment and our place in it. Clearly, situation awareness is what we know about the situation or our mental model of the situation.

In 1995, Endsley proposed a three-step model to represent the processes sustaining the building of situation awareness. The resulting mental model, from the perception, comprehension and projection cognitive processes, is inherently subjective, based on integrating acquired information with our own personal structural and situational factors. The quality of our SA may be characterized by the degree to which it accurately reflects objective reality. In other words, having good SA means that there is a high level of fidelity between the content of our mental model of the situation and the situation itself. Even if it is not an easy task, finding ways to measure the "goodness of fit" between the subjective interpretation of the reality (SA content) and the reality itself has been one of the main interest of the human factor engineering community.

The task of measuring SA is further complicated in situations for which a team of individuals have to work together (virtually or physically). According to Breton and Rousseau (2003), there are multiple political or social reasons to constitute to execute a task. Most C2 environments are now involving multiple nations with their respective set of rules, doctrines, cultures, etc. However, from a task execution standpoint, a team is required when the amount of information to be processed is too important for one individual or when the information types require different expertises owned by different individuals. It results that in both circumstances, team members have only access to a specific portion of the information.

From communication and coordination activities, they can build a common understanding of the situation or a shared awareness of the situation. This shared SA is a critical element for the execution of the task. According to Perla et al. (2000), TSA is accomplished through three functions: i) build individual situational awareness; ii) share their individual situational awareness; iii) develop the group's shared situational awareness. The degree to which team members can achieve these three functions is extremely important for developing effective team performance.

In the three functions proposed by Perla et al., clearly the first one can be evaluated with the individual SA measures. The measure can be used on the different team members in order to define the quality of their respective SA. The second function can be measured by evaluating communication and coordination activities between team members. The measure used may not be related to SA. It may further address the quality and the number of communications between team members. The third function proposed by Perla et al., may be seen as the most complicated one to evaluate. It is a

challenge to clearly define what it is shared within a team. Researches should be directed towards developing effective tools focusing particularly on this third function.

## 7. Conclusion

---

The advent of the information era at the end of last century created a new challenge for persons involved in C2 environments. In most situations, humans are overwhelmed by large number of information sources. Difficulties arise from the need to select the appropriate information at the right time. Therefore, it is critical to identify what information is required and how and when it should be made available to the human operator. During the last decade, SA theories and models have been developed in an attempt to provide answers to these questions. Concurrently with theory development, the need for measurement tools and protocols has emerged.

One of the most important obstacles to the development of measurement tools has been the multiplicity of SA definitions. In a literature review, Breton and Rousseau (2001) have listed 26 different definitions associated with the SA concept. In fact, in attempting to define SA as either the process or the state, many researchers in the field have asserted their own conceptualization of SA. One reason for the abundance of SA definitions is that SA must be dependent on the situation in which the SA task is performed. Consequently, many SA definitions, some more distinct than others, have been put forward to define SA in a given context (see Breton and Rousseau, 2001). Endsley (1995b) asserted that, although there are numerous SA definitions, most are not applicable across different task domains. Endsley proposed what is probably the most quoted “general” definitions of SA. “Situation awareness is 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”.

The dependency of SA on the situation in which the task is performed led to the development of an even larger number of SA measures as reported in this document. These different measures are needed to deal with the specific constraints inherent in different situations and environments. Another reason for the multitude of SA measures is the coupling of the measurement process with the definition process. To measure a concept reliably, one must define it appropriately. Researchers in the field of SA are faced with the problem of choosing the right definition and the most appropriate measure. This situation led Breton and Rousseau (2003) to propose a method for selecting a measurement tool based on the objective of the measurement process and a set of factors related to the situation in which the SA task is performed and to the person performing the task.

In addition to the need to select the right information from a daunting number of information sources, another consequence of the expanding information load in C2 environments is the need of teamwork to execute C2 activities. Teams are required when the number of information sources is too high to be processed by one individual, when the types of information sources to be processed require different kinds of expertise possessed by different individuals. In the latter situation, the team must be multi-disciplinary. Now, more and more C2 situations involve coalitions of different countries with different rules, doctrines, cultures, etc. The need for measuring SA in team situations stems for the increasingly frequent employment of teams and

coalitions in C2 environments. According to Pigeau and McCann (2000), one key aspect of C2 is the propagation of the commander's intent among subordinates. This aspect underscores the importance of sharing information between team members and, measuring the quality of the shared SA process and the percentage of information shared. For Perla et al. (2000), team situation awareness is the result of three functions. First, team members must build their own SA. Second, they must share the content of their respective SA, and third, they must develop the group's shared SA.

Measuring TSA may be the equivalent of evaluating the quality of the execution of these three functions (measuring the team situation assessment process) or the quality of the shared SA content (measuring TSA). In both cases, measuring the first function can be done with the use of individual SA measurement tools, but then it may suffer from the same limitations. The choice of a definition for SA and the consequent choice of a tool are determining factors in the quality of the evaluation process. The measurement of the second function is related to the evaluation of communication and coordination activities. Communication and coordination are processes not restricted to SA. Consequently, measures usually taken for these processes (i.e., frequency of messages and task overlapping) can be used to evaluate the second function in the Perla et al, proposition. The third function is probably the most complex to evaluate. Measuring the *shared* SA content among team members raises a number of questions, given that sharedness as such is an applicable concept. Schwartz (1990) states that each individual flight crew member possesses a level of situation awareness, but that TSA could not be calculated simply as the sum of the situation awareness achieved by each crew member. Further research is required to clarify the concept of shared SA and avoid the potential pitfalls of so-called shared cognition (see Rousseau, Tremblay and Breton, 2004).

This document reports a list of measures for evaluating individual and team SA. In addition to the different SA measures, their psychometric properties and their applicability to C2 environments are presented. The evaluation of psychometric properties has been hampered by the fact that most validation studies are too specific to the often-unique situation in which SA was measured and that reports are not always available in the open literature. The analysis conducted in this study reveals two needs that must be addressed in the future. First, psychometric evaluations of SA tools should be made available in the open literature. Second, further research is required to define and operationalize the concept of TSA.

## References

---

- Adams, S. (1998). Practical considerations for measuring Situational Awareness. Proceedings for the Third Annual Symposium and Exhibition on Situational Awareness in the Tactical Air Environment, 157-164.
- Adams, M. J., Tenney, Y. J., and Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors*, 37(1), 85-104
- Andersen, H. H. K., and Hauland, G. (2000). Mesuring team situation awareness of reactor operators during normal operation: A technical pilot study. Paper presented at the Proceeding of the Fourth International Symposium on Aviation Psychology, Savannah, Georgia.
- Banbury, S., Hoermann, H. J., Soll, H., and Dudfield, H. J. (in review). Development and evaluation of novel measures to assess the effectiveness of commercial airline pilot situation awareness training. Manuscript under review [*International Journal of Aviation Psychology*].
- Banbury, S., and Howes, A. (2001). Development of generic methodologies for the evaluation of collaborative technologies. DERA Technical Report (Contract CU005-2927).
- Baker, D. P., and Salas, E. (1992). Principles for measuring teamwork skills. *Human Factors*, 34, 469-475.
- Baker, C. V., Salas, E., Cannon-Bowers, J. A. and Spector, P. (1992). The effects of inter-positional uncertainty and workload on teamwork and task performance. Paper presented at the 7th annual conference of the society for industrial and organizational psychology, Montreal, Canada.
- Bolman, L. (1979). Aviation accidents and the theory of the situation. In G. E. Cooper, M. D. White, and J. K. Lauder (Eds.), *Resource management on the flight deck: Proceedings of a NASA Industry workshop* (pp. 31-58). Moffett Field, CA: NASA Ames Research Center.
- Bolstad, C. A., and Endsley, M. R., (2000). The effect of task load and shared displays on team situation awareness. Proceedings of the 14<sup>th</sup> Triennial Congress of the International Ergonomics Association and the 44<sup>th</sup> Annual Meeting of the Human Factors and Ergonomics Society, 2000.
- Brannick, M. T., and Prince, C. (1997). An overview of team performance measurement. In M. Brannick, E. Salas and C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods and applications*. Hillsdale, NJ: Erlbaum.

- Brannick, M. T., Prince, A., Prince, C., and Salas, E. (1995). The measurement of team process. *Human Factors*, 37, 641-651.
- Brannick, M. T., Salas, E., and Price, C. (1997). *Team Performance Assessment and Measurement. Theory, Methods, and Applications*. Lawrence Erlbaum Associates.
- Breton, R., and Rousseau, R. (2001). *Situation Awareness : A review of the concept and its measurement*. (DRDC Valcartier TR 2001-220. Defence RandD Canada - Valcartier.
- Breton, R. and Rousseau R., (2003). *The Analysis of Team Decision Making Architectures. Human Factors of Decision Making in Complex Systems*, Dunblane, Scotland, 8 – 12 septembre, 2003.
- Bryant, D., J., (2003). *Critique, Explore, Compare, and Adapt (CECA) : A new model for command decision-making*. Defence Research and Development Canada-Toronto, DRDC TR 2003-105, 61 pages.
- Cannon-Bowers, J. A., Salas, E., and Blickensderfer, E. (1999). *Toward an understanding of shared cognition*. Unpublished manuscript.
- Cannon-Bowers, J. A., and Salas, E. (1997). *A framework for developing team performance measures in training*. In M. Brannick, E. Salas and C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods and applications*. Hillsdale, NJ: Erlbaum.
- Carretta, T. R., Perry, D. C., and Ree, M. J. (1996). *Prediction of situational awareness in F-15 pilots*. *The International Journal of Aviation Psychology*, 6, 21-41.
- Cooper, M., Shiflett, S., Korotin, A. L., and Fleishman, E. A. (1984). *Command and control teams: Techniques for assessing team performance (final report)*. Wright-Patterson AFB, OH: Logistics and Human Performance Division. In Dickinson, T. L., and McIntyre, R. M. (1997). *A conceptual framework for teamwork measurement*. In M. T. Brannick, E. Salas and C. Prince (Eds.). *Team performance assessment and measurement: Theory, methods and applications*. Hillsdale, NJ: Erlbaum.
- Croft, D. G., Banbury, S. P., Aymeric, B., and Dudfield, H. J. (2000). *Enhanced Safety through Situation Awareness Integration in training*.
- Dickinson, T. L., and McIntyre, R. M. (1997). *A conceptual framework for teamwork measurement*. In M. T. Brannick, E. Salas and C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods and applications*. Hillsdale, NJ: Erlbaum.
- Durso, F. T., and Gronlund, P. (1999). *Situation awareness*. In F. T. Durso, et al (Eds.), *Handbook of applied cognition*. New York, NY: John Wiley and Sons.

- Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., Nikolic, D., and Manning, C.A. (1998). Situation awareness as a predictor of performance for en route air traffic controllers. *Air Traffic Control Quarterly*, 6(1), 1-20.
- Endsley, M. R. (1987). SAGAT: A methodology for the measurement of situation awareness (NORDOC 87-83). Hawthorne, CA: Northrop.
- Endsley, M. R. (1988a). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society 32nd Annual Meeting*, Santa Monica, CA, 97-101.
- Endsley, M. R. (1988b). Situation awareness global assessment technique (SAGAT). *Proceedings of the National Aerospace and Electronics Conference (NAECON)* (pp. 789-795). New York: IEEE
- Endsley, M. R. (1990). Predictive utility of an objective measure of situation awareness. *Proceedings of the Human Factors Society 34th Annual Meeting* (Vol. 1, pp. 41-45). Santa Monica, CA: Human Factors Society.
- Endsley, M. R. (1995a). Measurements of situation awareness in dynamic systems. *Human Factors*, 37(1), 65-84.
- Endsley, M. R. (1995b). Towards a theory of Situation Awareness. *Human Factors*, 37, 32-64.
- Endsley, M. R. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, (pp. 82-86). Santa Monica, CA: Human Factors and Ergonomics Society.
- Endsley, M. R. (2000a). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley, and D. J. Garland (Eds), *Situation Awareness Analysis and Measurement* (pp. 147-173). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Endsley, M. R. (2000b). Situation awareness in aircraft maintenance teams. *International Journal of Industrial Ergonomics* 26(2): 301-325.
- Endsley, M. R., and Garland, D. J., (2000). The effect of task load and shared displays on team situation awareness. *Proceedings of the 14<sup>th</sup> Triennial Congress of the International Ergonomics Association and the 44<sup>th</sup> Annual Meeting of the Human Factors and Ergonomics Society, 2000.*
- Endsley, M. R., Holder, L. D., Leibrecht, B. C., Garland, D. J., Wamper, R. L., and Matthews, M. D. (2000). Modeling and measuring situation awareness in the infantry operational environment. U. S. Army Research Institute for the Behavior and Social Sciences. Infantry Forces Research Unit.

- Endsley, M. R., and Rodgers, M. D. (1998). Distribution of attention, situation awareness, and workload in a passive air traffic control task: Implications for operational errors and automation. *Air Traffic Control Quarterly*, 6(1), 21-44.
- Endsley, M. R., Sollenberger, R., and Stein, E. (2000). Situation awareness: A comparison of measures. Proceedings of the first human performance, situation awareness, and automation conference. Savannah, Georgia.
- Ericsson, K. A., and Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251.
- Ericsson, K. A., and Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: The MIT Press.
- ESSAI (2001). WP2 Identification of factors affecting Situation Awareness and Crisis Management on the flight deck. Workpackage Report. ESSAI/QINETIQ/WPR/WP2/3.0. EC DG-TREN. Contract No.: 2000-GRD1-10450.
- Fadok, D.S., Boyd, J., and Warden, J., (1995). Air power's quest for strategic paralysis. Maxwell Air Force Base AL: Air University Press, (AD-A291621).
- Fink, A. A., and Major, D. A. (2000). Measuring team situation awareness: A comparison of three techniques. Paper presented at the Proceeding of the Fourth International Symposium on Aviation Psychology, Savannah, Georgia.
- Flin, R., and Martin, L. (2001) Behavioural markers for Crew Resource Management. *International Journal of Aviation Psychology*, 11, 95-118.
- Foy, J. L., and McGuinness, B. (2000). Implications of cockpit automation for crew situational awareness. Proceedings of the first human performance, situation awareness, and automation conference. Savannah, Georgia: October 2000.
- Fowlkes, J., Lane, N., Salas, E., Franz, T., and Oser, R. (1994) Improving the measurement of team performance: The TARGET's methodology. *Military Psychology*, 6, 47-61.
- Gawron, V. J. (2000). *Human performance measures handbook*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Glanzer, M., Glaser, R., and Klaus, D. J. (1956). The team performance record: An aid for team analysis and team training (AD-123 615). Washington, DC: Office of Naval Research, Psychological Science Division.
- Glickman, A. S., Zimmer, S., Montero, R. C., Guerette, P. J., Campbell, W. J., Morgan, B. B., and Salas, E. (1987). The evolution of teamwork skills: An empirical assessment with implications for training (NTSC Tech. Rep. No. TR-87-016). Orlando, FL: Naval Training Systems Center.

- Gravell, M., and Schopper, A. W. (1994). Measures and methods for characterizing operator performance in nuclear power plant control rooms: workshop findings and recommendations. US Nuclear Regulatory Commission Report #NRC FIN W6104.
- Gronlund, S. D., Ohrt, D. D., Dougherty, M. R. P., Perry, J. L., and Manning, C. A. (1998) Role of memory in air traffic control. *Journal of Experimental Psychology: Applied*, 4, 263-280.
- Gugerty, L. J. (1997). Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied*, 3(1), 42-66.
- Gutwin, C., and Greenberg, S. (2004). The importance of team cognition in distributed collaboration. In E. Salas and S. M. Fiore (Eds.) *Team Cognition: Understanding the factors that drive process and performance*, pp. 177-201, Washington: APA Press.
- Hartel, C. E. J., Smith, K., and Prince, C. (1991). Defining aircrew coordination: Searching mishaps for meaning. Paper presented at the Sixth International Symposium on Aviation Psychology, Columbus, OH.
- Hauland, G. (2002). Measuring team situation awareness in training of en route air traffic control. Technical report Riso-R-1343(EN). Riso National Laboratory, Denmark.
- Helmreich, R. L., and Foushee, H. C. (1993). Why crew resource management? Empirical and theoretical bases of human factors training in aviation. In E. L. Weiner, B. G. Kanki, and R. C. Helmreich (Eds.), *Cockpit resource management* (pp. 3-45). San Diego, CA: Academic.
- Hillburn, B. (2000, October). Measuring the right stuff: A comprehensive approach to aerospace SA assessment. Proceedings of the first human performance, situation awareness, and automation conference. Savannah, Georgia.
- Horswill, M. S., and McKenna, F. P., (2004). Drivers' hazard perception ability: Situation awareness on the Road. In S. Banbury, and S. Tremblay (Eds), *A cognitive approach to situation awareness: theory and application*, (pp. 155-175). Ashgate Publishing Limited, England.
- Hoffman, R. R. (1997). The elicitation of expert knowledge in domains involving perceptual skill, mental workload, and situation awareness. Unpublished manuscript.
- Jones, D. G., and Endsley, M. R. (2000a). Examining the validity of real-time probes as a metric of situation awareness. Proceedings of the 14th Triennial Congress of the International Ergonomics Association and the 44th Annual Meeting of the Human Factors and Ergonomics Society.

- Jones, D. G., and Endsley, M. R. (2000b, October). Can real-time probes provide a valid measure of situation awareness? In D. B. Kaber, and M. R. Endsley, Proceedings of the first human performance, situation awareness, and automation conference: Human performance, situation awareness and automation: User-centered design for the new millennium. Savannah, Georgia.
- Kaber, D. B., and Endsley, M. R. (1998). Team situation awareness for process control safety and performance. *Process Safety Progress*, 17, 43-48.
- Larson, C. E., and LaFasto, F. M. J. (1989). *Teamwork: What must go right / What can go wrong*. Newbury Park, CA: Sage.
- Matthews, M. D., Pleban, R. J., Endsley, M.R., and Strater, L.D. (2000). Measures of infantry situation awareness for a virtual mout environment. In D. B. Kaber, and M. R. Endsley, Proceedings of the first human performance, situation awareness, and automation conference: Human performance, situation awareness and automation: User-centered design for the new millennium. Savannah, Georgia.
- McGuinness, B., and Ebbage, L. (2001). Assessing human factors in command and control: Workload and situational awareness metrics. BAe Sowerby Research Centre technical report.
- McGuinness, B., and Foy, J. L. (1999). CARS (Crew Awareness Rating Scale): A subjective tool for situational awareness assessment. BAe Sowerby Research Centre technical report JS-14261.
- McGuinness, B., and Foy, J. L. (2000). A subjective measure of SA: The Crew Awareness Rating Scale (CARS). Proceedings of the first human performance, situation awareness, and automation conference. Savannah, Georgia: October 2000.
- McIntyre, R. M., Salas, E. Morgan, B., and Glickman, A. (1989). Team research in the 80's: Lessons learned. Technical report. Orlando, FL: Naval Training Systems Center.
- Morgan, B. B., Glickman, A. S., Woodward, E. A., Blaiwes, A. S., and Salas, E. (1986). Measurement of team behaviors in a navy environment (Tech report NTSC TR-860914). Orlando, FL: Naval Training Systems Center.
- Mosier, K. L., and Chidestser, T. R. (1991). Situation assessment and Situation Awareness in a team setting. In Y. Queinnec and F. Daniellou (Eds.), *Designing for Everyone* (pp. 798-800). London: Taylor and Francis.
- Muniz, E. J., Salas, E., Stout, R. J., and Bowers, C. A. (1999). The validation of a team Situational Awareness Measure. Proceedings for the Third Annual Symposium and Exhibition on Situation Awareness in the Tactical Air Environment (pp. 183-190). Patuxent River, MD: Naval Air Warfare Center Aircraft Division.

- Muniz, E. J., Stout, R. J., Bowers, C. A., and Salas, E. (1997). A methodology for measuring team situation awareness. Paper presented at the 18th Annual Industrial/Organizational Psychology/Organizational Behavior Graduate Student Conference, Roanoke, VA.
- O'Hare, D. (1997). Cognitive ability determinants of elite pilot performance. *Human Factors*, 39, 540-552.
- O'Hare, D., and O'Brien, K. (2000). Individual differences in situation awareness and training for complex tasks. In A. R. Lowe and B. J. Hayward (Eds.), *Aviation resource management: Vol. 2* (pp. 369-377). Aldershot: Ashgate.
- Paradis, S., Treurniet, W., and Roy, J. (1998). Environment perception process in maritime command and control. *Sensor Data Fusion and Integration of the Human Element*. DREV.
- Parks, D., Johnson, B., Picard, G., Dubois, S., MacLean, J., Ouellet, J., and Craig, M., (1997). Maritime C3I Way Ahead Working Group (MC3IWA WG) Final Report, DREV-TM-9715.
- Payne, J. W., Bettman, J. R., and Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY: Cambridge University Press.
- Perla, P. P., Markowitz, M., Nofi, A., and Weuve, C., (2000). *Gaming and shared situation awareness*. Alexandria, VI: Centre for Naval Analyses.
- Pew, R. W. (1995). The state of Situation Awareness measurement: Circa 1995. *Proceeding of the International Conference on Experimental Analysis and Measurement of Situation Awareness*, Embry-Riddle Aeronautical University Press, FL.
- Pew, R.W. (2000). The state of situation awareness measurement: Heading toward the next century. In M. R. Endsley, and D. J. Garland (Eds.), *Situation Awareness Analysis and measurement* (pp. 33-47). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pigeau, R., and McCann, C. (2000). Re-defining Command and Control. In C. McCann and R. Pigeau (Eds), *The human in command: Exploring the modern military experience* (pp. 163-184). New York, NY: Kluwer Academic/Plenum Publishers.
- Prince, C., and Salas, E. (1989). Aircrew performance, coordination and skill development. In D. Daniels, E. Salas, and D. M. Kotick (Eds.). *Independent research and independent exploratory developments (IE/IED) programs: Annual Report FY88* (Tech. Report NTSC 89-009). Orlando, FL: Naval Training Systems Center.

- Prince, C., and Salas, E. (1993). Training and research for teamwork in the military aircrew. In E. L. Wiener, B. G. Khaki, and R. C. Helmreich (Eds.), *Cockpit resource management* (pp. 337-366). San Diego, CA: Academic.
- Prince, D., Salas, E., and Brannick, M. (1999). Situation awareness: what do we know now that the 'buzz' has gone? In D. Harris (Ed.) *Engineering psychology and cognitive ergonomics: Vol. 3* (pp. 216-221).
- Ricks, W. R. T., Jonsson, J. E., and Barry, J. S. (1996). Managing approach plate information study (MAPLIST): An information requirements analysis of approach chart use (NASA Technical Paper 3561). Hampton, VA: NASA Langley Research Center.
- Robertson, M. M., and Endsley, M. (1994). The role crew resource management (CRM) in achieving team situational awareness. Paper presented at the Western European Association for Aviation Psychology 21st Conference, Dublin, Ireland.
- Rousseau, R. and Breton, R., (2004).\_ The M-OODA: A Model Incorporating Control Functions And Teamwork In The OODA Loop. Proceedings of the 2004 Command and Control Research and Technology Symposium, San Diego, U.S.A., June 2004.
- Rousseau, R., Tremblay, S., and Breton, R., (2004). Defining and modeling situation awareness, a critical review. In S. Banbury, and S. Tremblay (Eds), *A cognitive approach to situation awareness: theory and application*, (pp. 3-21). Ashgate Publishing Limited, England.
- Roscoe, S. N., Corl, L., and LaRoche, J. (2001). *Predicting human performance*. Saint Laurent, Qc: Helio Press Inc.
- Salas, E., Prince, C., Baker, D. P. and Shrestha, E. (1995). Situation Awareness in team performance: Implications for measurement and training. *Human Factors*, 37, 123-136.
- Sarter, N. B., and Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *International Journal of Aviation Psychology*, 1, 45-57.
- Sarter, N. B., and Woods, D. D. (1995). How in the world did we ever get in that mode? Mode error and awareness in supervisory control. *Human Factors*, 37(1), 5-19.
- Schvaneveldt, R. W. (1990). *Pathfinder associative networks: Studies in knowledge organisation*. Norwood, NJ: Ablex.
- Schwartz, D. (1990). *Training for situational awareness*. Houston, TX: Flight Safety International.

- Selcon, S. J., and Taylor, R. M. (1990). Evaluation of the situational awareness rating technique (SART) as a tool for aircrew system design. In *Situational awareness in aerospace operations (AGARD-CP-478)* (pp. 5/1-5/8). Neuilly-sur-Seine, France: NATO-AGARD.
- Selcon, S. J., and Taylor, R. M. (1991). Decision support and situational awareness. In R. M. Taylor (Ed.), *Situational awareness in dynamic systems (IAM Report 708)*. Farnborough, UK: Royal Air Force Institute of Aviation Medicine.
- Selcon, S. J., Taylor, R. M., and Koritsas, E. (1991). Workload or situational awareness?: TLX vs. SART for aerospace systems design evaluation. In *Proceedings of the Human Factors Society, 35th Annual Meeting (Vol. 1, pp. 62-66)*. Santa Monica, CA.
- Shebilske, W. L., Goettl, B. P., and Garland, D. J., (2000). Situation awareness, automaticity and training. In M. R. Endsley, and D. J. Garland (Eds.), *Situation Awareness Analysis and measurement* (pp. 303-324). Mahwah, NJ: Lawrence Erlbaum Associates.
- Shrestha, E., Prince, C., Baker, D. P., and Salas, E. (1995). Understanding situation awareness: Concepts, methods and training. In W. B. Rouse (Ed.), *Human / technology interaction in complex systems: Vol 7.* (pp. 45-83). Connecticut: JAI Press.
- Stout, R., Cannon-Bowers, J. A., and Salas, E. (1994). The role of shared mental models in developing shared situational awareness. In R. D. Gilson, D. J. Garland, and J. M. Koonce (Eds.), *Situational awareness in complex systems* (pp. 17-23). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.
- Strater, L. D., Endsley, M. R., Pleban, R. J., and Matthews, M. D. (2001). Measures of platoon leader situation awareness in virtual decision making exercises. (Research Report 1770). Alexandria, VA: Army Research Institute.
- Sullivan, C., and Blackman, H. S. (1991). Insights into pilot Situation Awareness using verbal protocol analysis. *Proceeding of the Human Factors Society 35th Annual Meeting*, (pp.57-61). Santa Monica, CA: Human Factors and Ergonomics Society.
- 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)* (pp. 3/1-3/17). Neuilly-sur-Seine, France: NATO-AGARD.
- Taylor, R. M., and Selcon, S. J. (1991). Subjective measurement of situational awareness. In R. M. Taylor (Ed.) *Situational awareness in dynamic systems (IAM Report 708)*. Farnborough, UK: Royal Air Force Institute of Aviation Medicine.

- Taylor, R. M., Selcon, S. J., and Swinden, A. D. (1995). Measurement of situational awareness and performance: a unitary SART index predicts performance on a simulated ATC task. Proceedings of the 21st Conference of the European Association for Aviation Psychology. Chapter 41.
- Vidulich, M. A. (1989). The use of judgements matrices in subjective workload assessment: The subjective workload dominance (SWORD) technique. Proceeding of the Human Factors Society 33 rd Annual Meeting (pp.1406-1410). Santa Monica, CA: Human Factors and Ergonomics Society.
- Vidulich, M. A. (2000). Testing the sensitivity of situation awareness metrics in interface evaluations. In M. R. Endsley, and D. J. Garland (Eds), Situation Awareness Analysis and Measurement (pp. 227-246). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Vidulich, M. A., and Hughes, E. R. (1991). Testing a subjective metric of situation awareness. Proceedings of the Human Factors Society 35th Annual Meeting (pp. 1307-1311). Santa Monica, CA: Human Factor Society.
- Vidulich, M. A., Stratton, M., Crabtree, M., and Wilson, G. (1994). Performance-based and physiological measures of situational awareness. Aviation, Space, and Environmental Medicine, 65, (5 Suppl.), A7-A12.
- Waag, W. L., and Houck, M. R. (1994). Tools for assessing situational awareness in an operational fighter environment. Aviation, Space, and Environmental Medicine, 65(5 Suppl.), A13-A19.
- Wagner, M., and Simon, R. (1990). Development of measures of crew coordination. (Contract DAHC35-89-D-0030).
- Wellens, A. R. (1993). Group situation awareness and distributed decision making from military to civilian applications. In N. J. Castellan (Ed.). Individual and group decision making: Current issues (pp. 267-291). Hillsdale, NJ: Erlbaum.

## **List of symbols/abbreviations/acronyms/initialisms**

---

DND	Department of National Defence
C2	Command and Control
SS	Support System
SA	Situation Awareness (state) or Situation Assessment (process)
DM	Decision Making
OODA	Observe-Orient-Decide-Act
TSA	Team Situation Awareness
EPOG	Eye Point of Gaze
SIREQ-TD	Soldier Information Requirements-Technology Demonstrator
DRDC	Defence Research and Development Canada
EEG	Electroencephalographic Activity
SABARS	Situation Awareness Behaviorally Anchored Rating Scale
SAGAT-MOUT	Situation Awareness Global Assessment Technique – Military Operations in Urbanized Terrain
PSAQ	Participant Situation Awareness Questionnaire
SAGAT	Situation Awareness Global Assessment Technique
ATC	Air Traffic Controller
SPAM	Situation Present Assessment Method
SART	Situation Awareness Rating Technique
SALIENT	Situation Awareness Linked Indicators Adapted to Novel Tasks
SARS	Situation Awareness Rating Scale
SA-SWORD	Situation Awareness - Subjective Workload Dominance Technique
CLSA	China Lake Situation Awareness

IMPSA	Implicit Situation Awareness toolkit
SASRF	Situational Awareness Supervisory Rating Form
CARS	Crew Awareness Rating Scale
TCAS	Traffic Collision Avoidance System
CSA	Crew Situational Awareness
IMA	Integrative Methods Approach
TARGETs	Targeted Acceptable Responses to Generated Events or Tasks
SAW-IM	Situation Awareness and Workload in-Flight Measure
FASA	Factors Affecting Situation Awareness
IPTE	Information Propagation in Team Environments

## Distribution list

---

### Internal Distribution by Centre:

- 1- Director General
- 3- Document Library
- 1- Head/Decision Support Systems Section
- 1- Head/Information and Knowledge Management
- 1- Head/System of Systems Section
- 1- Head/Optronic Surveillance Section
- 1- LCol Pierre Lefebvre (DJCP CTG C4ISR Team Leader)
- 1- Mr Claude Roy (SDA)
- 1- Dr Richard Breton (author)
- 1- Dr Pierre Valin
- 1- Mr Jean Berger
- 1- Dr Patrick Maupin
- 1- Mrs Micheline Bélanger
- 1- Dr Anne-Claire Bourry Brisset
- 1- Mr François Rhéaume
- 1- Mr Stéphane Paradis
- 1- Mr Martin Blanchette
- 1- Dr Hengameh Irandoust
- 1- Dr Abder Benaskeur
- 1- Dr Abdeslem Bouktouta
- 1- Dr Mohammad Allouche
- 1- Dr Anne-Laure Jouselme
- 1- Mrs Catherine Daigle
- 1- Dr Adel Guitouni
- 1- Mr Eric Dorion
- 1- Mr Abder Sahi
- 1- Mr Jean Roy
- 1- Mrs Régine Lecocq
- 1- Mr Denis Gouin
- 1- LCol Michel Gareau
- 1- LCdr Elizabeth Woodlife
- 1- LCol Jean-Guy Lévesque
- 1- Dr Marielle Mokhtari

External Distribution by DRDKIM:

NDHQ  
101 Colonel By Drive  
Ottawa ON K1A 0K2

- 1- Director – R&D Knowledge and Information Management (PDF file)
- 1- Col Rick Williams (Associate DG RDP)
- 1- Director – Directorate Science and Technology (Air)
  - Directorate Science and Technology (Land)
  - 1- attn: Mr Craig Maskell
  - 1- attn: Maj David Walker
  - Directorate Science and Technology (Maritime)
  - 1- attn: Dr Pierre Lavoie (DSTM)
  - 1- attn: Mr Greg Walker (DSTM 3)
  - 1- attn: Mrs Laura Ozimek (DSTM 5)
  - Directorate Science and Technology (Policy)
  - 1- attn: Mr Clément Laforce
  - 1- attn: Mr Robert Webb
  - 1- attn: Mrs Lisa Willner
- 1- Director – Directorate Science and Technology (Human Performance)
  - Directorate Science and Technology (C4ISR)
  - 1- attn: Dr Chris McMillan (DSTC4ISR)
  - 1- attn: Mr Laus Kollenberg (DSTC4ISR 3)
  - 1- attn: Mrs Donna Wood (DSTC4ISR 4)
  - 1- attn: Mrs Caroline Wilcox (DSTC4ISR 5)
  - Centre for Security Science (CSS)
  - 1- attn: Dr Anthony Ashley - (Director General)
  - 1- attn: Mrs Susan McIntyre - (Knowledge Manager)
  - 1- attn: Dr Walter Dyck - (DSTHP)
  - 1- attn: Director – CBRN Research and Technology Initiative (CRTI)
  - 1- attn: Director - Public Security Technical Program (PSTP)
  - DRDC Centre for Operational Research and Analysis (CORA)
  - 1- attn: Dr Kendall Wheaton
  - 1- attn: Mr Dave Connell
  - 1- attn: Mr Ron Funk
  - 1- attn: Mr Paul Comeau
  - 1- attn: Mr Paul Pulsifer (DGIP/DSTI)
- 1- Cdr Stephen R. Peters – CMS – DGMFD – DMRS2 (Maritime Security Section Head)
- 1- LCdr Jim Day – CMS – DGMFD – DMRS2 (Project Director MSOC)
- 1- LCol John Langen – DPDOIS (Project Director MSOC)
- 1- LCol Jacques Hamel – DLCSPM 7 (PMO ISTAR)
- 1- Capt (N) Darren Knight – CMP – CDA
- 1- Mr Paul Morin – DCDS – CDI – D INT IM
- 1- LCol Terrence Procyshyn – DJFC 4 (PD JIIFC)
- 1- LCol Craig Dewar – DJFC (D/PD JIIFC)
- 1- LCol Thomas Sullivan – ADM (IM) (PM JIIFC)

- 1- Mr Robert McCaulay – ADM (IM) (D/PM JIIFC)
- 1- LCol Phil Jourdeuil – ADM (IM) - DGIMSD- DIMSP
- 1- LCol Rick Johnston – VCDS – CFD – DJCP  
VCDS – DGSC – DKM
- 1- attn: LCol Christopher Blodgett
- 1- attn: Mrs Karen Lahaise
- 1- attn: Mr Mike Crowell
- 1- Cdr Haydn Edmundson – J3 Maritime – Support and Tasks – Canada Command HQ

DRDC Atlantic  
9 Grove Street  
Darmouth NS B2Y 3Z7

- 1- Dr Ross Graham – Director General - DRDC Atlantic (Scientific Advisor (Maritime))
- 1- Mr Jim L. Kennedy
- 1- Cdr Siegfried Richardson-Prager (SMO)
- 1- Mr Jim S. Kennedy (Thrust Leader “11h”)
- 1- Dr Mark McIntyre
- 1- Mr Dave Chapman
- 1- Mrs Liesa Lapinski
- 1- Mr William Campbell
- 1- Mr Mark Hazen

DRDC Toronto  
PO Box 2000  
1133 Sheppard Avenue  
Toronto ON M3M 3B9

- 1- Dr Ross Pigeau
- 1- Mrs Carol McCann
- 1- Mrs Sharon McFadden
- 1- Dr David Smith
- 1- Dr Justin Holland
- 1- Dr Joe Baransky
- 1- Dr Renee Chow
- 1- Dr David Bryant
- 1- Dr Keith Hendy

DRDC Ottawa  
3701 Carling Avenue  
Ottawa ON K1A 0Z4

- 1- Mr Chris Helleur
- 1- Mr Rick Brown
- 1- Mr Jack Pagotto
- 1- Mrs Barbara Waruszinsky

Canadian Forces Experimentation Centre (CFEC)  
3701 Carling Avenue  
Ottawa ON K1A 0Z4

1- Capt (N) Kevin Laing  
1- LCol Tom Gibbons  
1- LCol J.D. Graham  
1- LCdr Rob Elford  
1- Dr Phillip Farrell  
1- Dr Fred Lichacz

1 -Dr Sébastien Tremblay (co-author)  
École de Psychologie,  
Université Laval,  
Québec, Québec,  
Canada, G1K 7P4

UNCLASSIFIED  
SECURITY CLASSIFICATION OF FORM  
(Highest Classification of Title, Abstract, Keywords)

<b>DOCUMENT CONTROL DATA</b>		
1. ORIGINATOR (name and address) Richard Breton, Defence Research and Development Canada-Valcartier 2459 boul. Pie-XI Nord Québec, Qc, G3J 1X5	2. SECURITY CLASSIFICATION (Including special warning terms if applicable) Unclassified	
3. TITLE (Its classification should be indicated by the appropriate abbreviation (S, C, R or U) Measurement of individual and team situation awareness: A critical evaluation of the available metrics and tools and their applicability to Command and Control environments (U)		
4. AUTHORS (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.) Breton, Richard; Tremblay, Sébastien, Banbury, Simon		
5. DATE OF PUBLICATION (month and year) May 2007	6a. NO. OF PAGES 108	6b. NO. OF REFERENCES 121
7. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. Give the inclusive dates when a specific reporting period is covered.) Technical Report		
8. SPONSORING ACTIVITY (name and address)		
9a. PROJECT OR GRANT NO. (Please specify whether project or grant) 11b12	9b. CONTRACT NO.	
10a. ORIGINATOR'S DOCUMENT NUMBER TR 2003-283	10b. OTHER DOCUMENT NOS  N/A	
11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)		
<input checked="" type="checkbox"/> Unlimited distribution <input type="checkbox"/> Restricted to contractors in approved countries (specify) <input type="checkbox"/> Restricted to Canadian contractors (with need-to-know) <input type="checkbox"/> Restricted to Government (with need-to-know) <input type="checkbox"/> Restricted to Defense departments <input type="checkbox"/> Others		
12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)		

UNCLASSIFIED  
SECURITY CLASSIFICATION OF FORM  
(Highest Classification of Title, Abstract, Keywords)

UNCLASSIFIED  
SECURITY CLASSIFICATION OF FORM  
(Highest Classification of Title, Abstract, Keywords)

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

The objectives with this document are to provide a critical evaluation of the psychometric properties of the available individual and team situation awareness (SA) measurement tools and techniques and to evaluate the applicability of the SA measurement tools found in the literature to C2 environments. This report is based on an exhaustive literature survey of SA measurement and a search for tools and techniques that have been used to measure individual and team SA. The emphasis is on the psychometric properties of the tools, such as reliability, validity and sensitivity. The survey reveals a wide variety of tools, techniques and metrics but very little information about their psychometric properties. In many cases validity and reliability data are rather inconclusive and incomplete with rather small samples. Nevertheless some well-established measures of SA such as SAGAT and SART have proved to be useful in relation to the evaluation of new interface or design and training programs. One key problem with SA measurement is the wide variety of measures for a large number of situations. There are nearly as many different measurement techniques as there are different situations in which an operator is at work and each situation has its own sets of procedures, task requirements and demands. According to the distribution of the SA measurement tools among the C2 environments, the SA concept can be measured in every situation. In some occasions, a slight modification of an existing measure can be brought to adapt that tool to another situation. (U)

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Situation Awareness, Situation Assessment, Measurement, Measurement tool, Assessment Method, Human Performance, Teamwork, Team Performance, Team Shared Awareness, Team Shared Assessment

UNCLASSIFIED  
SECURITY CLASSIFICATION OF FORM  
(Highest Classification of Title, Abstract, Keywords)



## **Defence R&D Canada**

Canada's Leader in Defence  
and National Security  
Science and Technology

## **R & D pour la défense Canada**

Chef de file au Canada en matière  
de science et de technologie pour  
la défense et la sécurité nationale



[WWW.drdc-rddc.gc.ca](http://WWW.drdc-rddc.gc.ca)

