Towards measurement of Situation Awareness in Virtual Reality Simulator

Sunniva Veie [(1)], Salman Nazir [(1)]

[(1)] *Training and Assessment Research Group (TARG), Department of Maritime*

Operations. Faculty of Technology, Natural Sciences & Maritime Sciences,

University of South-Eastern Norway, Horten, 3184, Norway, Salman.Nazir@usn.no

Keywords: virtual reality, training, situation awareness.

ABSTRACT: The increase in global trade has made seaports more complex both in traffic and scale. It is often a challenge for the ports to handle the fast-paced cargo movement while maintaining safe and efficient operation. The use of high-tech equipment and computerized cargo management has alleviated the challenge to a greater degree and introduced novel complexities in different levels of the system for human operators. Basic training or mere familiarization is not always enough for the human operators to master the sophisticated task in ports where millions of dollars are at stake. Computer simulation coupled with newly emerging Virtual Reality (VR) simulation has evolved as tools for training. VR is both inexpensive and mobile compared to traditional high-fidelity fixed simulators. More importantly, Head-mounted Virtual Reality (HMD-VR) can simulate and provide the "feel" of the workplace surroundings in an immersive environment. As VR has been deployed for port training, situation awareness (SA) measurement in VR training has become an issue. Measuring SA in an activity may indicate the level of control an individual has obtained over the situation and may explain why the person acts in a specific way. The performance of trainees may also be greatly affected by differing levels of SA. Thus, SA measurement can provide inputs to the training methods which could also help in the evaluation and optimization of those methods (Nazir et al., 2015). SA can be measured through various ways in traditional training, such as performance measures, freeze probe techniques and self-rating techniques. However, the practicality of these methods in HMD VR training remains questionable. This study compared the usefulness of existing SA measurement methods in HMD VR simulator based on assessments by four subject matter experts (SME) and shows that existing methods are potentially applicable. Then a new procedure based on a combination of existing methods and slight modifications of these was developed to optimize the SA measurement in HMD VR simulator.

INTRODUCTION

Virtual reality is an emerging simulation technology and an important part of the development of "smart ports" that can simulate immersive training scenarios. VR simulators, from low immersive 2 DOF (Degrees of freedom) simulators with flat-screen to immersive 6 DOF simulators (Agostino G et al, 2013) and now HMD VR simulators, are being used in marine port equipment training activities.

Measurements of SA level in trainees are deemed useful to evaluate and optimize training procedures and the output of training (Endsley, 1995b). All published studies related to VR and SA mainly focus on the impact of VR on SA in training (Read, J. M., & Saleem, J. J, 2017; R. M. S. Clifford et al, 2018) or using VR as a tool to enhance or measure SA (Minji Choi et al, 2020), rather than SA measurement methods in VR.

Situation awareness has been a heated topic of discussion for researchers throughout the years (Stanton, Salmon, Walker, Baber, & Jenkins, 2005) since SA measurement methods have been developed for traditional training scenarios. However, unlike traditional scenarios, a HMD VR simulator enhances immersion which may affect the ability of the participant to obtain an overview of a situation. Therefore, this paper aims to identify an optimal method for the measurement of SA in VR environments by evaluation and combination of existing methods.

SITUATION AWARENESS

Early stages of research on SA were conducted during and after the first world war and was chiefly focused on optimizing the performances of aviation teams, by analyzing how well the aviation team members assessed and controlled specific situations. It was found that increasing SA in-flight processes were crucial for the high performance of the teams. Although SA was mainly used in aviation, it permeated soon to other domains. It was applied in the analysis of large system operations like nuclear power plants, in smaller less complex (but still high security) systems like police and firefighter processes, and even in every-day activities like driving (Endsley, 1995b). To understand the extension of situation awareness, it is necessary to know how situation awareness affects the participant's behavior. Endsley (1995) divides SA into three levels: Perception, comprehension and projection, corresponding to awareness of the environment, ability to understand and predict the effects of an action. Endsley (1995) further develops a SA model,

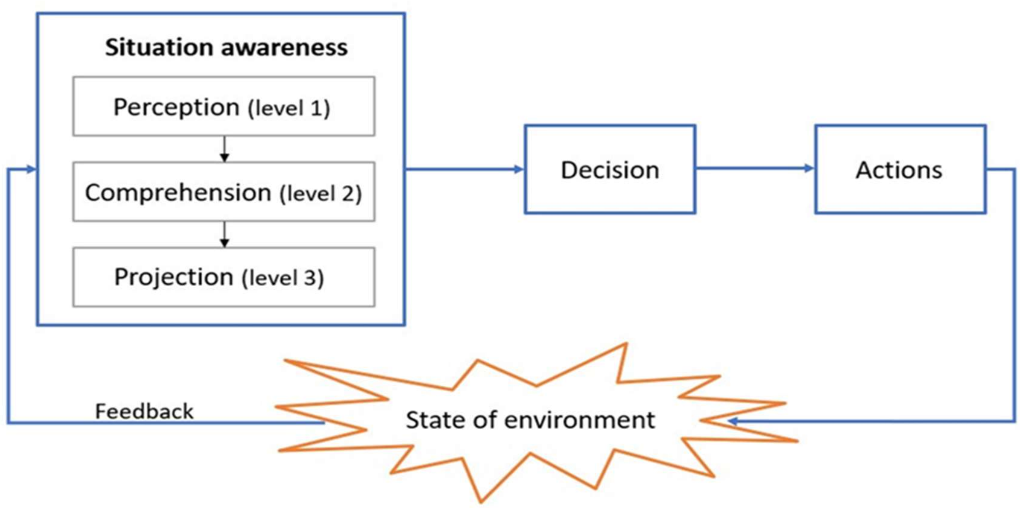

Figure 1: Modification of Endsley (1995) Model of Situation Awareness

The model delineates perception as the initial level, where the elements of the situation are perceived. The elements are then further investigated in level two, where the understanding of the elements is focused. Level one and level two is then used in level three as a base for projection of future status of the elements and the situation. The three levels comprise the SA that is used in a decision-making process resulting in a decided action to reach a desired goal state of the environment. When the action is decided, it is initiated with an objective of high performance. The actions taken could influence the state of the environment, and preferably fix the instability or problem which was identified in the SA levels. The new state of the environment is then investigated with the three levels of SA, and if the goal state is not reached, the whole process is initiated again. Several factors, like information processing and system capacity, will interfere and provide input to the stages of the process, affecting the output.

SA MEASUREMENT IN TRAINING

As SA affects the performance of real-life activities and work, it also affects the performance and outcome of training (Endsley, 1995b). Poor SA in training generally leads to less optimal training outcome (Nazir et al., 2012), by focusing on wrong elements or ignoring important elements in the training module. Measurement is important to benchmark the quality of the training standards and for the assessment of the method itself.

There are several SA measurement techniques and methods. Investigating the **SA requirements** is the initial step for many SA analysis (Stanton et al., 2005). In this method, SMEs focus on the environment where the SA will be measured and attempt to identify elements of SA within this environment. **Freeze probe techniques** are much-used methods of SA measurement. The methods consist of administered freezes conducted during the training scenarios, where the simulation stops and the screens go blank and the computer or simulator initiates relevant SA queries for tasks completed by the participant in the simulation so far (Stanton et al., 2005). **Real-time probe techniques** are similar to the freeze probe techniques, except the real-time probes do not stop the simulation and thus avoid much of the intrusiveness (Stanton et al., 2005). These techniques introduce probes with queries during the simulation. **Self-rating techniques** are measurement methods were the participants themselves rate their perceived SA performance based on subjective measurements (Stanton et al., 2005). The participants are provided with a chart of several dimensions of SA, and are instructed to rate these dimensions from their perceived SA performance. These ratings are conducted after the training has been completed, measuring the performance of the whole training period. **Observer-rating techniques** are measurement techniques where an observer monitors the training sequence and rates the SA performance based on his/her observations. These methods often involve SMEs observing and rating the participants, to obtain the best information on the important aspects of the procedures. The measurements are based on rating different SA skills or behaviors performed by the participants during the training exercises (Stanton et al., 2005). **Performance measures** are then used to evaluate the performance of the participants in the training scenario. These methods attempt to measure how participants perform in different situations, and how they react to changes or manipulated errors in the system. **Similarity index** is introduced as a method of measuring team members responses or performance by comparing them to the best-informed team member or to a team average (Sætrevik & Eid, 2013). Considering intrusiveness, cost, and the result's reliability, each of the above methods has advantages and disadvantages. These methods can be concluded in Table-1,

Table 1: SA Measurement Methods (adapted from Stanton et al., 2005)

Techniques	*Methods*	*Features*
SA requirements analysis	GDTA	Used to clarify elements for further analysis. Useful in creating queries for measurements, need SMEs to include in the analyzing, and resource consuming.
Freeze probe	SAGAT	A direct measure of participants SA in all three SA levels, and it is easy to use, however, this method is High intrusiveness and time consuming to develop measurements.
Real-time probe	SPAM	A direct measure of participants SA and less intrusive than SAGAT, but it's time-consuming and high burden on SMEs, also the construct validity is very limited.
	SASHA	Similar to SPAM, but with automated measurement. SASHA_L for real-time measurement and SASHA_Q for queries after training.
Self-rating	SART	Conducted after test completion (long-/short term memory of participants), but it's easy to implement with low cost and low intrusiveness. Uses 10 dimensions to measure SA.
	SARS	Similar to SART but uses 7 dimensions to measure SA.
Observer-rating	SABARS	Rating based on SMEs observation, low intrusiveness, but requires qualified SMEs, also the construct validity is low.
Performance measures	Global measures	Automated analyzing by computer.
	External task measures	Real-time measurement but highly intrusive
	Imbedded task measures	Non-intrusive and high validity of measures, but hard to determine measures. High SA in one area might lower the SA in another.
Similarity Index		Low intrusiveness. But only applicable in team measurement.

MEASURE SA IN VR

From the first simulator created in the late 1930s to advanced full-scale simulators, and then VR simulators nowadays, training simulators have been developed with increasing accuracy regarding training quality and immersion. SA measurement methods are changing with simulators' development. For example, in full-scale simulators, measurement methods used to measure SA in real-life situations are relatively appropriate, as the simulator scenarios and equipment resemble real-life activities. Figure 2 illustrates how initial SA measurement methods have been perceived as applicable to "traditional simulators", and how there might be a need for new methods for measuring SA in the virtual environment within the next generation of training simulators.

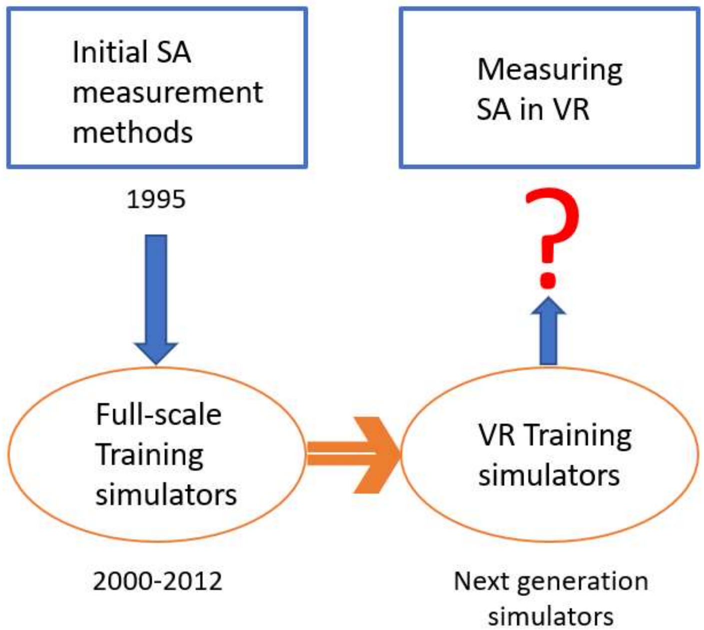

Figure 2: Development of new simulators requiring new measurement methods

As VR simulation training still is a relatively novel approach in training. It could be important to investigate the SA of the participants involved to optimize training output in the head-mounted VR simulation. Additionally, the SA of the participants in a virtual environment may be different from that of other training environments, as the enhanced immersion may affect the ability of the participant to obtain an overview of a situation. To identify if SA is better in VR than in other training environments, measurements enabling a comparison of SA within different training methods are needed.

Several human factors, like attention and working memory, affect the situation awareness of people in real life- and training situations alike. Lacking attention may lower the SA, as elements of the environment will be overlooked (Endsley, 1995b). Training in VR simulators enables an immersive experience that may be perceived as transferable to real-life training. If the VR simulator manages to raise the attention of the user, it may raise the perception and in the end the SA. Thus, measurements of SA in VR may also provide information validating if virtual reality enhances the SA in training.

To evaluate the usefulness of SA measurement methods in HMD VR, four SMEs received a table of the different measurement methods, and a short description of the VR training environment and the measurements methods. The SMEs were then asked to provide their comments on the usability of these methods in measuring SA in VR. They were also asked to rate the different methods from one to eleven, eleven being the most applicable method and one the least. Figure 3 shows the summary of ratings.

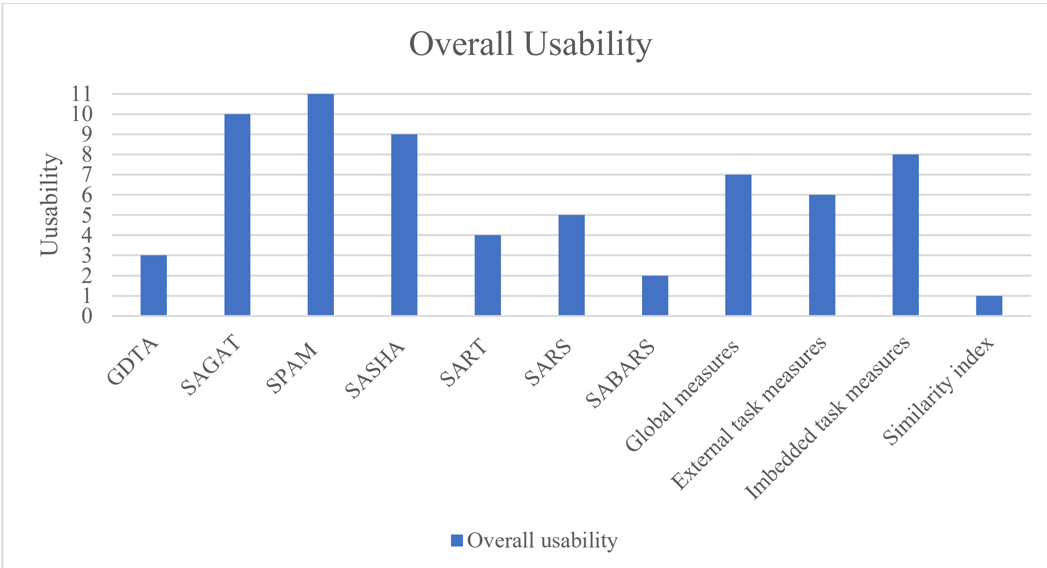

Figure 3: Overall Usability in VR

Figure 4 shows an overview of SME ratings, which have been divided into three levels: High, medium and low score. The intervals of the scores are as follows: High score: 11-8, medium score:7-5, low score: 4-1. As one of the SMEs did not include comments on the similarity index, this method only includes three ratings.

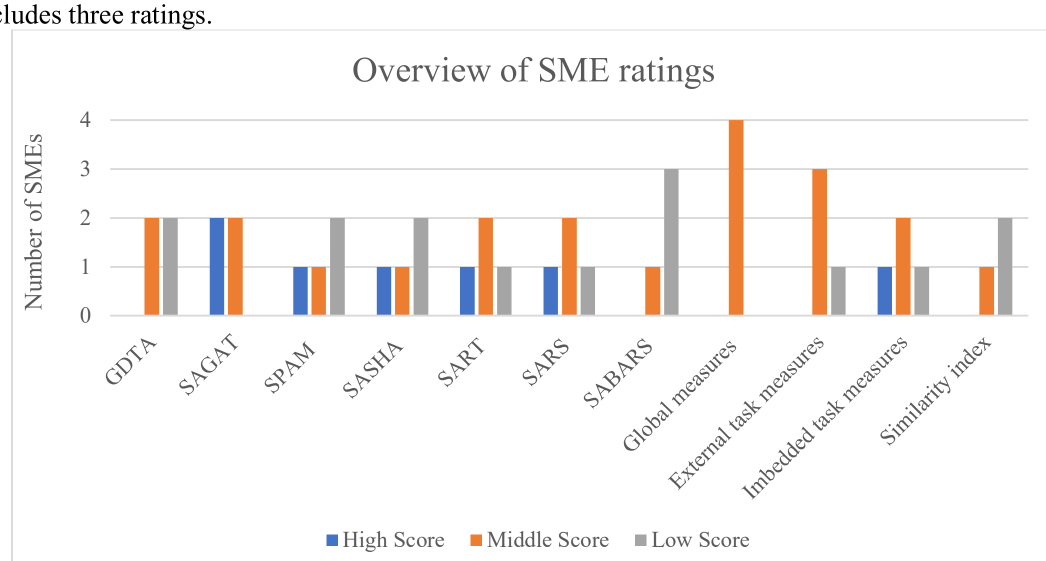

Figure 4: Overview of SME ratings

As the above diagram divides the SME rating results into three groups, based on each SMEs response, it might be misleading. The methods with more high ratings can be viewed as poor because of one or two low ratings. Thus, their score on the above figure seem lower than their total score. Even though SAGAT here looks to be rated the highest, SPAM actually received the total highest scores (see figure 3). Thus, the diagram presents an overview of the SMEs ratings, but not necessarily an exact picture of the results.

Most of the SMEs state that the SA requirements analysis may not be classified as a tool measuring SA in task performance, but that it is useful for creating further measurements queries for other measurement methods. Both the freeze-probe technique and the real-probe technique was generally recognized as methods of high usability in VR simulator training, but some of the experts also pointed out the intrusiveness of these methods may affect the outcome. About fifty percent of the SMEs considered the validity of self-rating methods low, as the subjective view of SA may not be very reliable or correct. Some of the SMEs argue that self-rating is a highly pragmatic method which can provide important information regarding the experience of the participant, which may enable improvements of the training scenario and a more complete evaluation of the performance. Thus, self-rating methods were generally deemed useful but limited, and it was mentioned that the method should be used as a supplement to other methods. The SME opinions on the observer-rating methods included in this study were rather negative, as the view will be fairly limited from the outside of the head-mounted gear used in VR simulators. Like self-rating methods, SMEs suggested using observer-rating methods as a complement to other methods. Most SMEs argued the similarity index method low usability in measuring SA in VR simulators, since this method is not suitable for individual training, and most HMD VR training is conducted individually. The response regarding the usability of the performance measures methods was generally positive. The VR environment can be controlled and designed to test the participant automatically throughout the exercises, performance measures may provide highly valid information on the participant's SA. The negative aspects presented by the SMEs is that quality of programming and implementation of the queries within the VR scenario may affect the perceived immersion and performance of the participant.

The SMEs view of the methods generally indicates that performance measures, real-time probe techniques and freeze probe techniques are most applicable. Other methods, like self-rating and observer-rating techniques, was argued to provide important data on performance and thus recommended to be used as supplement methods of measuring SA in VR training simulators.

OPTIMAL METHOD

Based on SME ratings and comments, a new procedure with a mix of these methods with slight modifications was identified, it is a combination of performance measures, real-time probes, self-evaluation and observation (see Figure 5). This new procedure also collects SA performance data during and after the training session, aims to eliminate the limitations of the existing old procedures and provide a superior and robust assessment system. SA requirements analysis was introduced as an analysis of the environment and provide information for creating measurements queries for other measurement methods. The method employed for probing will be similar to SASHA Method, where probing during the exercise and a post-exercise questionnaire will be used to assess the trainee's performance and experience. Global measures, such as score and time, and embedded task measures measuring secondary task performance will be included. The equipment will be programmed to conduct an automatic comparison of individual performances against other participants. Thus, providing "updated score chart" of trainee's performances, enabling comparison of performance to the performance of the "top performer". A simplified observation technique is also included, to assess the external problems such as equipment discomfort. With HMD VR, it will be easier to use devices such as eye-tracking equipment. Eye-tracking can provide a direct visual measurement of attention (Lo & Meijer, 2013), so researchers can obtain data on which elements are subject to most attention. Self-rating methods are also employed to provide information pertaining to the training experience.

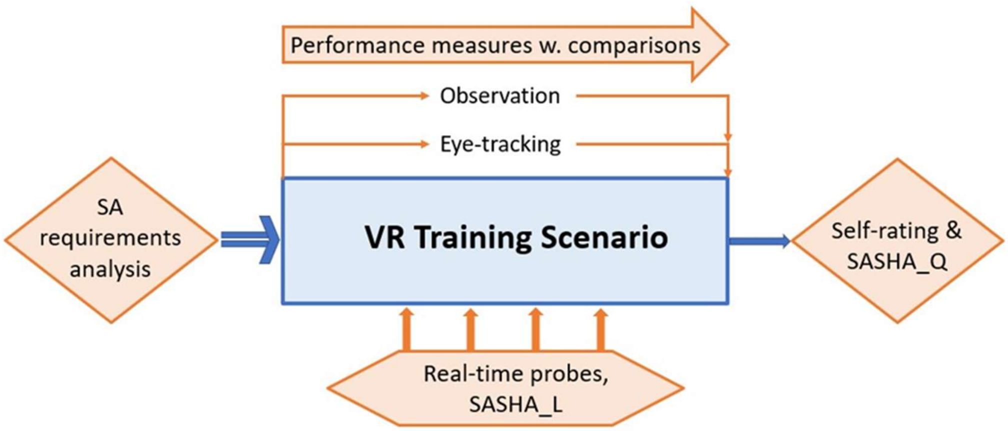

Figure 5: Model of optimal SA measurement procedure

CONCLUSION

This study of situation awareness measurement methods was conducted to evaluate the usability of different methods when applied to VR training simulators. Eleven methods have been identified and evaluated, both by the authors and by four SMEs. The applicability of these methods in a virtual environment was deemed variable. However, most methods were argued to be useful even if only as a supplement to other measurement methods.

An optimal method/procedure based on a combination of different methods, some slightly modified, have been developed. However, this method was not tested, the study lacks empirical evidence validating the method. In addition, the number of SMEs evaluating the usability of the methods was relatively low (four). Including evaluations from more SMEs would increase the strength of the results.

Further research could steer towards testing of suggested method in comparison with other emerging new methods. Moreover, the effectiveness of utilizing them in measuring SA in virtual environments such as the head-mounted VR simulators could be explored. Exploring the effect of enhanced immersion on SA may reveal the usability of VR training simulators, which may contribute to the enhanced use of such training methods. Differing methods targeting specific training scenarios could be developed and tested, for example, VR training for port crane operations and forklift operations may require different measurement methods.

ACKNOWLEDGEMENTS

The second author would like to thank the Research Council of Norway, and Kongsberg Digital AS for funding the project *Innovating Maritime Training Simulators using Virtual and Augmented Reality (InnoTraining)* (project number: 269424)

REFERENCES

Agostino G. Bruzzone, Francesco Longo. (2013). 3D simulation as training tool in container terminals: The TRAINPORTS simulator. *Journal of Manufacturing Systems*, 32(1), 82-98.

Endsley, M. R. (1995a). Measurement of Situation Awareness in Dynamic Systems. *Human Factors the Journal of the Human Factors and Ergonomics Society, 37*(1), 65-84.

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

Lo, J. C., & Meijer, S. A. (2013). Situation Awareness Measurement Techniques for Gaming simulations: An Overview and Application for Railway Traffic Controllers. Paper presented at the IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, San Diego USA.

Minji Choi, Seungjun Ahn, JoonOh Seo. (2020). VR-Based investigation of forklift operator situation awareness for preventing collision accidents. *Accident Analysis & Prevention,* 136.

Nazir, S., Colombo, S., & Manca, D. (2012). The role of situation awareness for the operators of process industry. Chemical Engineering Transactions, 26.

Nazir, S., Sorensen, L. J., Øvergård, K. I., & Manca, D. (2015). Impact of training methods on Distributed Situation Awareness of industrial operators. Safety science, 73, 136-145.

R. M. S. Clifford, H. Khan, S. Hoermann, M. Billinghurst & R. W. Lindeman. (2018) The Effect of Immersive Displays on Situation Awareness in Virtual Environments for Aerial Firefighting Air Attack Supervisor Training. IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Reutlingen, 2018, pp. 1-2.

Read, J. M., & Saleem, J. J. (2017). Task Performance and Situation Awareness with a Virtual Reality Head-Mounted Display. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61(1), 2105–2109.

Sætrevik, B., & Eid, J. (2013). The "Similarity Index" as an Indicator of Shared Mental Models and Situation Awareness in Field Studies. *Journal of Cognitive Engineering and Decision Making, XX*(X), 1-18.

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. P. (2005). *Human Factors Methods, A Practical Guide for Engineering and Design*. England: Ashgate Publishing Limited.