Authoring virtual simulations to measure situation awareness and understanding

Stella Polikarpus
The Estonian Academy of Security Sciences
stella.polikarpus@sisekaitse.ee

Tobias Ley
Tallinn University
tley@tlu.ee

HansHazebroek
Nederlands Instituut Publieke Veiligheid
hans.hazebroek@nipv.nl

Graham Edgar
University of Gloucestershire
gedgar@glos.ac.uk

Geoffrey Sallis
University of Gloucestershire
gsallis@glos.ac.uk

Steven Baker
University of Gloucestershire
sbaker1@glos.ac.uk

Anna Figueras Masip
Insitut de Seguretat Publica de Catalunya
afigueras@gencat.cat

ABSTRACT

Measuring situation awareness (SA) and situation understanding (SU) is an important topic for Command & Control research. Virtual simulations (VS) have been proposed as a suitable method for measuring SA/U, but, there is little research into how to build scenarios for VS so that SA/U can be measured reliably. In this study, we used two different VS scenarios and the Quantitative Analysis of Situation Awareness (QASA) method to measure, and provide feedback on, actual and perceived SA/U. Two VS scenarios were tested in Estonia with 36 trainees. The results of the different scenarios were compared to establish whether the scenario storylines and authoring process resulted in differences in SA or SU. We conclude that reliable assessment results were produced with both authoring processes, and further suggest that the Collaborative Authoring Process Model for Virtual Simulations (CAPM) be used for VS creation.

Keywords

situation awareness (SA), situation understanding (SU), FireFront, The Collaborative Authoring Process Model (CAPM) for Virtual Simulations (VS), actual situation awareness (ASA), actual situation understanding (ASU), perceived situation awareness (PSA), perceived situation understanding (PSU), Quantitative Analysis of Situation Awareness (QASA).

INTRODUCTION

Effective performance in safety-critical situations requires both good situation awareness (SA) and situation understanding (SU). In dynamic and naturalistic decision-making situations, such as firefighting incidents, **actual situation awareness** (ASA) indicates how aware participants are of relevant aspects of the situation (FireFront Project, 2021). ASA is an evaluation of the participants' phenotype schema (a knowledge structure representing the current situation) obtained by comparing the individual's knowledge of the situation to the 'ground truth' (Edgar et al., 2018). For **perceived situation awareness** (PSA), participants subjectively indicate, how good they thought their SA was (FireFront Project, 2021). **Actual situation understanding** (ASU) reflects the individual's ability to use information available to them while combining it with their knowledge and experience (a genotype schema) to 'make sense' of the situation (Thoelen et al., 2020) - as compared to the objectively ideal assessment. **Perceived situation understanding** (PSU) indicates how good participants thought their SU was (FireFront Project, 2021).

This study presents data from the FireFront project (Thoelen et al., 2020) that is a further development of a previous project, FireMind (Arendtsen et al., 2016). FireFront has developed to measure actual and perceived SA (ASA/PSA), as well as actual and perceived SU (ASU/PSU). The aim of the project is to support the training and self-awareness (via automated feedback) of dynamic decision-making for incident commanders in fire and rescue operations. VS based videos are produced and then statements about these videos are presented, which participants are asked to respond to. Measures of SA are calculated in FireFront using the well-validated QASA (Quantitative Analysis of Situation Awareness) method (Thoelen et al., 2020) that asks participants whether statements about the scenario are true or false. SU is assessed by asking participants to judge whether information is relevant, or irrelevant, to the successful completion of the task.

Simulations are widely used for the measurement of SA (Endsley, 1995) and VS are seen as a valuable tool for assessment of firefighters' SA. Building effective scenarios is, however, a resource-heavy operation. One of the scenarios (the 'baseline-scenario' - BS) reported in this paper was built for the FireFront project by experts from six different countries using an iterative process to develop the scenario and the measures within it. Although this is a robust method for scenario building, it took a considerable amount of time and effort. Moreover, it is not feasible to employ a similarly resource intensive process of authoring scenarios in practical settings of training organizations where hundreds incident commanders are trained using VS. Furthermore, there is also evidence that it is beneficial to involve trainers in the development of the scenarios, rather than using ready-made or expert generated scenarios. This is because the former ones can be adjusted to local settings and trainers themselves gain knowledge about SA and its assessment. Therefore, the efficient construction of different VS scenarios is important for both practical and scientific reasons.

For that incentive, we created a second scenario (we call "national scenario"-NS) following the Collaborative Authoring Process Model (CAPM), a methodology for involving trainers into the creation of VS scenarios (Polikarpus & Ley, 2021). CAPM represents an efficient and tightly-specified method for generating VS scenarios. The generalizability of the CAPM needs, however, to be established as it is unclear whether CAPM can be used for building scenarios for different SA/U measurement method.

The current paper has the following aims: to extend the applicability of CAPM (Polikarpus et al., 2020) to scenarios that provide computer generated feedback on ASA, PSA, ASU and PSU - and to establish that the scenarios developed using the CAPM are functionally equivalent to those developed using the more time-consuming method in FireFront.

To address these aims we ran two scenarios with the same participants. We then collected ASA, PSA, ASU, and PSU data using both VS scenarios in a classroom setting under researcher supervision. To test the CAPM applicability for scenario creation we compared data collected for the CAPM scenario (NS) with the expert-generated BS. In the event that no significant differences between ASA/U are found and that there is a high correlation between the scores, we can assume that the two ways of creating scenarios give equivalent results and therefore the practical use of CAPM is justified in training organizations.

BUILDING VIRTUAL SIMULATION SCENARIOS

As stated above to be able to measure SA/U VS scenarios are needed. The following subsections give overview about VS scenario creation, one for NS and one for BS.

BUILDING THE NATIONAL SCENARIO USING THE COLLABORATIVE AUTHORING PROCESS MODEL (CAPM)

To be able to measure ASA/U, the incident commander needs to be presented with a realistic and immersive scenario with which they can feel they are involved (Polikarpus et al., 2020). The CAPM was developed to be used to assess incident commanders' dynamic decision-making skills by means of VS and it is practically applicable in fire service training organizations as it involves trainers in the building process (Polikarpus & Ley, 2021). Table 1 presents the main steps of CAPM that were used to build the NS. The storyline is that there is a fire in a car repair shop, one unconscious person inside, and one injured casualty standing outside. In the burning room are gas cylinders and waste oil.

Table 1. The FireFront Collaborative Authoring Process Model for Virtual Simulation (Polikarpus & Ley, 2021)

Steps	Sub-activities	Producer activities	Scriptwriter activities
1.Establishing the conceptual basis for measuring SA (e.g., Effective Command, QASA, other SA measuring method). Use an Excel spreadsheet to develop storyline.	• Find a co-author • Write the storyline • Choose the software for visualizations	Works with visual aspects of scenario.	Drafts the storyline through dialog box and true/false; relevant/irrelevant statements.
2.Author the scenario using virtual reality or combine real-life media files.	• Create or combine illustrative materials • Find relevant situation elements • Add the videos and statements into FireFront	Combine situation elements into visualization.	Helps to create the situation elements visualization.
3.Technical testing: reviewed by two experts.	• Dynamic play-through of the scenario using FireFront • Discussion based on disagreed statements	Take part in discussions.	
4.Finalize the VS by modifying unclear statements.	• Amendments in video files if visualisation needs to be connected. • Adding missing elements or new statements or correcting statements' phrasing	Make amendments in visualization.	Make amendments in statements.
5.Pilot and evaluate with trainees.	• Statistical analyses of the group data.	Make the statistical analyses	

The NS was fully developed in XVR On-Scene (2020 version) software for incident commanders' assessment using the CAPM (Polikarpus & Ley, 2021) as follows:

Step 1. The statements required to measure ASA and ASU with QASA were inputted using an Excel spreadsheet that laid out the statements used, and the visual/auditory information in the scenario to which they referred.

Step 2. The scenario visualization in video format (including background noises) was created by the "producer" (third column in Table 1) using XVR-OS software from the instructor point of view. The "scriptwriter" (fourth column) used the incident commanders' point of view captured using software Camtasia 6 (TechSmith, 2021). Audio was recorded using computer microphone and added to the videos using Camtasia. Altogether, eight videos were created, the shortest 27 seconds and the longest 1 minute 20 seconds.

Step 3. Two further experts (independently of the authors) did a dynamic playthrough of the scenario. Each statement that was answered differently was discussed and reviewed. At the end of this testing phase, the suitability and phrasing of each test statement was agreed between the four experts.

Step 4. Changes were made to statements to ensure they reflected the scenario. This is less time consuming than changing the scenario to match the statements. New videos or pictures could, however, be added to the scenario if necessary to increase the quality of the scenario.

Step 5. Statistical data analysis (see Results section) was conducted after the scenario had been completed by trainees to eliminate or guide rewording of, any 'poor' (e.g. misunderstood) statements.

BUILDING THE BASELINE SCENARIO

The BS is a collaborative effort of several SA/U measurement international experts (Thoelen et al., 2020). It took around one year to agree and create the BS that suited incident commanders' ASA/U measurement in international context. The final scenario is a visualization of a road traffic accident between three cars in a small-town environment. The BS was built using XVR (On-Scene 2017 version). The statements to measure SA/U had to be adjusted for six countries to prevent the undue influence of national procedures and culture in interpreting the information. In the scenario there were three videos used to illustrate the accident site and changes in the situation.

The shortest video was 1 minute 47 seconds and the longest 6 minutes 20 seconds. The order of statements and phrasing of them had to be changed several times. While the video is played in FireFront the trainee sees it over the full screen (see figure 1).

Statements that could be answered true/false were used to measure ASA and statements that could be judged relevant/irrelevant were created to measure ASU. For example, statements like "*There was a motor bike involved in the accident*." (correct "False") or "*The police controlled vehicle access*." (correct "True") were used to measure ASA. The proportion of correct answers was used to calculate ASA. Statements like "*Knowing the sex of casualties involved is relevant;*" (correct "irrelevant") or "*At a road traffic collision (RTC) knowing if there are any environmental issues is relevant;*" (correct "relevant") were used to calculate the ASU.

PSA and PSU were assessed using the question "*How sure are you that the response you have given is correct?*" after every statement, and it could be answered on a 4-point scale: "Uncertain" (1); "Fairly uncertain" (2); "Fairly certain" (3) and "Certain" (4) (Edgar et al., 2018).

DATA COLLECTION

The FireFront application was tested by trainees using the BS and NS in 2021 in Estonia in groups of two to ten. (see Figure 1). The researcher gave a short presentation of the study and introduced the activities expected from trainees. Inside the application a consent form was "signed" by the trainees. Altogether, 36 trainees took the scenario using individual headsets and computers (Figure 1). After the trainees had done the scenarios raw data files from each computer were collected manually. All data analysis was done using the methods described in Edgar et al., (2018).

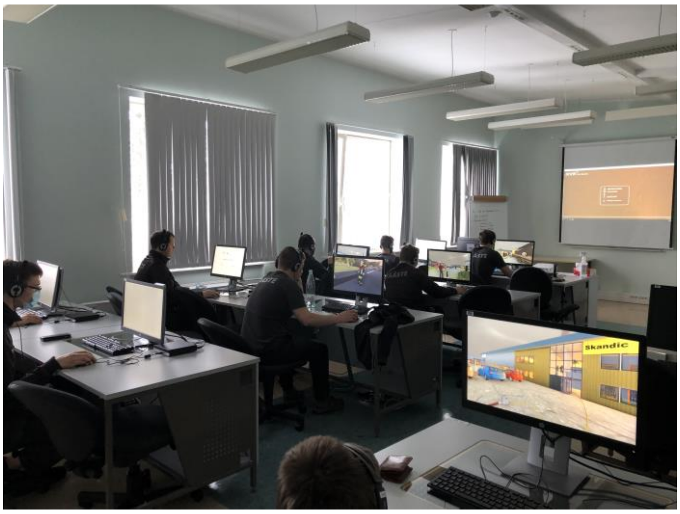

Figure 1. Testing the FireFront application in The Estonian Academy of Security Sciences

To answer **research question**: Does the CAPM generated scenario give similar results as experts created BS with the same ASA/U measurement method and group of trainees?" we calculated descriptive statistics and correlation analyses, with the hypothesis that performance on the two different scenarios would show significant correlations.

RESULTS

Answering RQ: Does the CAPM generated scenario give similar results than the BS for aspects of ASA/U?

The same participants (n=36) took the NS and BS using the same application version in a classroom setting. Figure 2 shows the mean values of ASA/U, and PSA/U for the scenarios. The standard deviation (SD) for the BS ASA was 12,3 and for the NS ASA SD=13,8 and the NS ASU SD=13,4, while the NS PSA SD=22,2 and PSU SD=27,1. The scores for all measures can run from -100 to +100. Note that a negative ASA or ASU indicates a misunderstanding of the situation on the part of the participant.

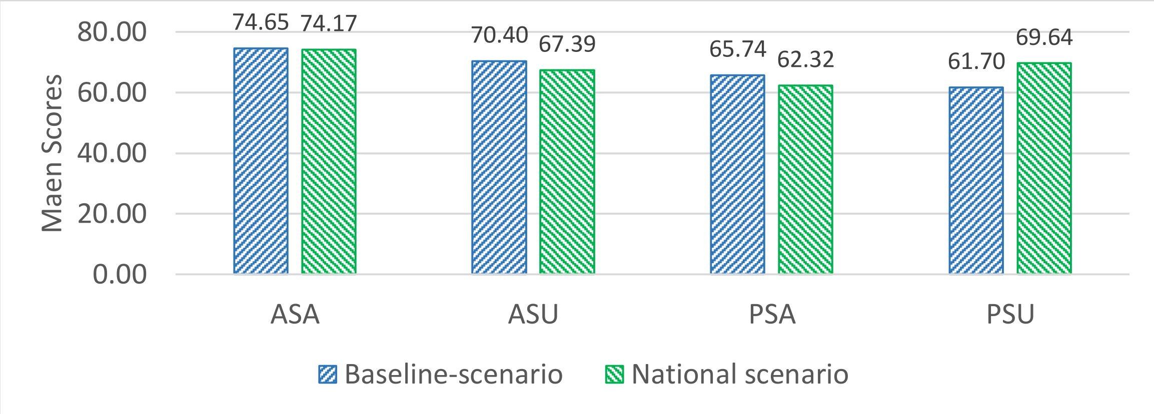

Figure 2. Comparison of baseline and national scenario results taken by same group of participants

To test whether the mean scores in the NS were comparable to the BS, we performed a t-test to compare the SA/U scores. There were no statistically significant differences between the two scenarios on any measure. We conclude that the two scenarios were comparable in the terms of the demands to SA/U VS based measurement using QASA.

A series of Pearson´s correlations were also conducted (see Table 2). All correlations indicated positive correlation between scores. There was a significant correlation between ASA BS and ASA NS scores (r=0,41; $p < 0.01$). There was a significant correlation between ASU for BS and NS (r=0,58; $p < 0.01$), indicating that performance was similar in the BS and NS scenarios.

We expected a positive correlation between PSA and PSU because it is measuring the general confidence level of incident commanders within a scenario (see Table 2 PSA BS and PSA NS r=0,76; $p < 0.01$; PSA BS and PSU BS r=0,77; $p < 0.01$; PSA BS and PSU NS r=0,68; $p < 0.01$ indicating medium correlation). ASA and ASU as well ASA and PSA/U did not correlate significantly within BS, which has also been found previously (Sallis et al., 2022).

Table 2. The correlation between baseline (BS) and national (NS) scenario results. Pearson's *r* is given in each cell and significant (p<0.05) correlations are shown in bold.

	ASA BS	ASA NS	ASU BS	ASU NS	PSA BS	PSA NS	PSU BS	PSU NS
ASA baseline (BS)	1,00							
ASA national (NS)	**0,41**	1,00						
ASU baseline (BS)	0,24	0,51	1,00					
ASU national (NS)	0,31	0,19	**0,58**	1,00				
PSA baseline (BS)	0,27	0,15	0,16	0,19	1,00			
PSA national (NS)	0,36	0,37	0,19	0,17	**0,76**	1,00		
PSU baseline (BS)	0,31	0,27	0,31	0,10	**0,77**	0,59	1,00	
PSU national (NS)	0,39	**0,41**	0,37	0,30	**0,68**	0,88	0,63	1,00

CONCLUSION

Efficient creation of VS scenarios is critical for operations of a training organization. While the BS was created using expert input and a tedious process of mapping SA/U requirements to the VS possibilities, CAPM was used for building the NS that involves trainers actively in the building process and is a more realistic method for practical purposes. The data suggest that CAPM can be used to generate a scenario (the NS) that produces similar results (for ASA/U) to the BS, that took a lot more time and resource to put together. These are promising results that need to be confirmed to be used in international context.

The scenarios described here were also some of the first to use QASA to assess SU as well as SA and this generated data that warrant further research; for example, that there was no significant correlation between ASA and ASU in either the NS or BS.

In summary, the implementation of QASA within the FireFront tool offers a robust method of measuring ASA/U and PSA/U. CAPM reduces the resources required to build valid scenarios for use within tools such as FireFront, and other tools. Providing robust scenarios that incorporate measures of SA/U that can both be used to assess firefighter performance, as well provide firefighters with awareness of these concepts, we believe are valuable additions to firefighter training.

Future studies will explore the use of CAPM to build a wider range of scenarios using a wider range of technologies (e.g. using video from body-cams) and also to investigate the use of the SA/U measures with other first responders in safety-critical situations.

ACKNOWLEDGMENTS

Co-funded by the Erasmus+ programme of the European Union. The FireFront team consisted of partners from Belgium, Denmark, Estonia, Netherlands, Spain and the UK. We appreciate the hard work of Frank Thoelen from Belgium for programming the FireFront application. Also, a big ‘thank you’ to all project members for developing the baseline-scenario.

REFERENCES

Edgar, G., Catherwood, D., Baker, S., Sallis, G., Bertels, M., Edgar, H. E., Nikolla, D., Buckle, S., Goodwin, C., & Whelan, A. (2018). Quantitative Analysis of Situation Awareness (QASA): modelling and measuring situation awareness using signal detection theory. *Ergonomics*, *61*(6), 762–777. https://doi.org/10.1080/00140139.2017.1420238

Endsley, M. R. (1995). Measurement of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 65–84. https://doi.org/10.1518/001872095779049499

FireFront Project. (2021). *FireFront Toolbox*. FireFront. https://firefront.eu/the-firefront-toolbox

Polikarpus, S., & Ley, T. (2021). Collaborative Authoring of Virtual Simulation Scenarios for Assessing Situational Awareness. In A. Adrot, R. Grace, K. Moore, & C. Zobel (Eds.), *Proceedings of the 18th ISCRAM Conference* (pp. 229–237). Blacksburg, VA, USA. https://www.drrm.fralinlifesci.vt.edu/iscram2021/files/ISCRAM2021Proceedings.pdf

Polikarpus, S., Ley, T., & Poom-Valickis, K. (2020). Developing the Situational Awareness of Incident Commanders: Evaluating a Training Programme using a Virtual Simulation. *Proceedings Estonian Academy of Security Sciences*, *19*, 195–226. https://digiriiul.sisekaitse.ee/bitstream/handle/123456789/2595/Proceedings_2020_WEB.PDF?sequence=5&isAllowed=y

Sallis, G., Catherwood, D., Edgar, G. K., Baker, S., & Brookes, D. (2022). Situation awareness and habitual or resting bias in high-pressure fire-incident training command decisions. *Fire Safety Journal*, *128*(103539). https://doi.org/10.1016/j.firesaf.2022.103539

TechSmith. (2021). *Camtasia 2020*. https://www.techsmith.com/camtasia-2020-press-release.html

Thoelen, F., Vastmans, J., Blom Andersen, N., Boehm, M., Holm, L., Arendtsen, B., Polikarpus, S., Taukar, M., Kütt, T., Fikke, R. C., Geertsema, T., Hazebroek, J. C., Tonnaer, C., Weewer, R., Figueras Masip, A., Fuste Castella, R., Catherwood, D., Baker, S., Brookes, D., … Walker, S. (2020). FireFront: A new tool to support training in Fireground Situation Awareness, Situation Understanding and Bias. *International Fire Professional*, *34*, 34–39.