Wright State University

CORE Scholar

International Symposium on Aviation Psychology - 2013

International Symposium on Aviation Psychology

2013

Assessing the Efficacy of Situation Awareness Probe Questions for Predicting Airtraffic- Management Performance

James D. Miles

Kim-Phuong L. Vu

Dan Chiappe

Thomas Strybel

Vernol Battiste

ASSESSING THE EFFICACY OF SITUATION AWARENESS PROBE QUESTIONS FOR PREDICTING AIR-TRAFFIC-MANAGEMENT PERFORMANCE

James D. Miles, Kim-Phuong L. Vu, Dan Chiappe, Thomas Strybel
Center for Human Factors in Advanced Aeronautics Technologies,
California State University Long Beach
Long Beach, CA USA
Vernol Battiste
NASA Ames Research Center, San Jose State University Foundation
Moffett Field, CA USA

We conducted an exploratory data analysis as a step toward modeling components of SA. It was based on data collected from situation awareness probe questions that were used in large-scale air traffic control simulations over 5 semesters of an ATC radar internship in the Center for Human Factors in Advanced Aeronautics Technologies (CHAAT). Three components of SA were generated by a principal component analysis that we label "Action Relevant," "Distance Relations," and "Low Priority." The analyses provide a data-driven scheme for categorizing probes related to SA that can be used in future evaluations of NextGen concepts and technologies.

Situation awareness (SA) is central to evaluating future airspace systems being considered by the FAA for implementation in the Next Generation Air Transportation System (NextGen). SA refers to the operator's understanding of the evolving situation that he or she is in for the purpose of projecting system states in the near future. The specifics of this definition have been debated in many recent reviews (e.g., Salmon, Stanton, Walker, and Jenkins, 2010; Chiappe, Strybel, and Vu, 2012; Jeannot, Kelly, and Thompson, 2003). Although airspace operators report having clear ideas of what SA means (e.g., D'Arcy and Rocco, 2001), Chiappe et al. (2012) showed that diverse theoretical perspectives on SA and its measurement are grounded in very different approaches to cognitive science. Endsley (1995a), for example, assumes that processing and representation all takes place in the conscious mind of operators. Recent distributed or situated conceptions of SA are instead based on a view of cognition that assumes operators offload task representation and computation to the external environment in order to limit use of internal processing resources (Salmon et al., 2010; Chiappe et al., 2012). According to situated SA theories, operators create partial representations of a situation that are constantly updated, and often internally store where to find information in the environment, rather than storing the information itself. SA, in this view, therefore exists in the interaction between the operator and his or her task environment.

These diverse concepts of SA have implications for how the construct is measured. Probe techniques are promising because they can assess an operator's awareness of specific information needed for adequate performance. Probe queries can be administered either offline or online. Endsley's Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1995b) is an offline method. In SAGAT, the operator is queried about task information when a scenario is frozen and the displays are blanked. SA is measured by the number of correct answers. This technique is consistent with the notion that knowledge of the task environment is stored internally because SA is determined by probing only the operator's working memory for task information. Evidence for the criterion validity of SAGAT has been reported (e.g., Endsley, 2000; Endsley, 1990b; Gronlund et al., 1998), and SAGAT has been used in a variety of settings such as air traffic control, aviation, and nuclear power plant operations. It is the most widely-used probe technique for situation awareness assessment to date. Online probe techniques such as Durso et al.'s (2004) Situation Present Assessment Method (SPAM) are consistent with a situated approach to SA. Operators are queried about task information in real time while performing their tasks with all displays and controls available for answering probe queries. Both, the number of correct responses and the latency of responses to queries are assumed to be indicators of SA, but only response latencies should be sensitive to whether information is offloaded to the environment. Evidence for the validity of this online probe technique has accumulated in recent years (e.g., Durso et al., 2004; 2006; Bacon et al., 2011; Strybel et al., 2013).

An important factor determining the effectiveness of all probe methods is the information contained in questions themselves. Yet, this factor has received little attention in the literature. For both offline and online probe techniques, it is recommended that the queries be developed in consultation with subject matter experts. Endsley et al. (2000) recommends that SAGAT queries be developed from a systematic Goal Based Task Analysis in which task goals, required information, and required SA for meeting the goals are identified. Durso et al. (2004) also recommend that queries be developed in conjunction with subject matter experts, but did not advocate a formal process of question development. However, for both probe methods, SA probe queries have been specific to the scenarios and variables manipulated, making it difficult to compare changes in SA across different simulations, experiments and operating concepts.

Work in our lab has been focused on developing and standardizing categories of probe-questions for assessing SA in current and future airspaces so that changes in awareness of specific information can be determined. We have examined several probe category schemes, but each has been found less than adequate for the purposes of comparison. Initially, we developed probe categories based on level of processing (recall, comprehension) and time frame (past, present, future) consistent with Endsley's conception of SA. Dao et al., (2009), for example, determined that pilot probe latencies for future questions were significantly longer than latencies to questions asking about present or past task information, suggesting that pilots were less aware of future information, especially for recall questions. Dao et al. noted that these results could be an artifact of the part-task simulation, however: pilots were queried about their airspace following individual conflict-resolution trials, and at that point the traffic in the scenario was frozen. Strybel et al. (2009) used the same information processing categories to measure ATCos' SA. They showed that categories based on task-specific information, such as conflicts, were more predictive of performance than information processing level or time frame. Subsequently, we developed probe categories based on the information required for different task components and operators; for example, conflicts, sector/aircraft status, command /communications, traffic and weather (e.g., Bacon et al., 2011Strybel et al., 2013). These categories were developed prior to a simulation based on consultations with subject matter experts and were more successful in predicting operator performance than the queries based on information processing categories. Recently, Morgan et al. (2012) categorized probe questions as part of a test of the situated SA approach. Morgan et al. hypothesized that questions based on high-priority task information would be answered more quickly than low-priority information, and that general-task questions would be answered more quickly than task-specific queries. This is because general information and high priority information should be stored in the head. Morgan et al. determined that probe latencies were faster for general information than for specific information, consistent with a situated situation awareness approach. However, they did not find an effect of task priority on probe latencies.

In summary, probe techniques are promising methods of assessing SA, yet their usefulness in comparing changes in SA across different simulations and comparing changes brought about by different concepts of operation are limited by a lack of standardized probe categories. Previous categorization schemes have been based on either theoretical assumptions of SA, or information relevant to specific task components. The problem of categorization is difficult because theory-based categories and task-specific categories are not mutually exclusive. For example, high-priority queries used in Morgan et al. (2012) were mostly questions on conflicts because safety is the highest priority of air traffic controllers. Similarly, future-oriented questions used in Strybel et al (2009) can be classified as conflict questions because detecting conflicts requires projection into the future.

In the current paper, we conducted an exploratory data analysis as the first step toward modeling components of SA. The analysis was based on data collected from probe questions related to ATC situational awareness that were used in simulations over 5 semesters of an air-traffic-control radar internship in the Center for Human Factors in Advanced Aeronautics Technologies (CHAAT). The advantage of such simulations is that they allow for the collection of real-time SA in a realistic ATC setting. In the current paper, we combined data from interns in these 5 simulations, which provided an appropriate N for conducting a principle component analysis.

Method

Participants

Data for this analysis were taken from 71 students enrolled in a radar internship course over five semesters (approximately 13 per semester) between 2010 and 2012. All students were enrolled in an FAA Collegiate Training Initiative (CTI) at the time of the course and had taken some courses in aviation sciences on FAA regulations and air traffic control operations. Students had little-or-no radar experience, however.

Training

The 16-week radar internship was designed to provide students training and practice managing traffic in a simulated en-route sector (ZID-91), using both current day manual skills and potential NextGen tools. The simulation software was Multi Aircraft Simulation System (MACS; Prevot et al., 2002). Students completed a 3-hour lab period and 2-hour lecture period each week. A retired, full performance Level (FPL) ATC taught the internship. Students learned to manage traffic with current day ATM techniques such as altitude, speed, vectoring, and structure, and current-day basic ATM phraseology. They also learned to manage traffic with simulated NextGen tools, specifically, integrated Data Comm, a conflict probe tool and a trial planner. The sequence in which current day and NextGen tools were introduced, as well as scenarios containing different mixtures of equipped and unequipped aircraft, varied somewhat between different semesters. At the end of each semester, however, all students had roughly equivalent practice with manual and NextGen skills, and had managed scenarios having 0% to 100% equipped aircraft.

Situation Awareness Assessment

At the midterm and end of the course, students were tested on three 40-50 minute scenarios differing in the percentage of equipped aircraft. These tests were intended to determine the effects of different training approaches, and were not aimed at evaluating their performance for classroom assessment (the instructor never saw the results). In the first semester, the equipage levels were 50% overall and varied within the scenarios between 25% and 75%. In the last four semesters, equipage levels were 0%, 50% and 100%. SA was also measured during these tests using an online probe technique. Students were instructed to respond to online probe questions about their airspace that were presented every 3 minutes beginning 4 minutes into the scenario. Probe questions were administered on an adjacent computer with touch-input. Each question began with a "ready for question" prompt, and the participant was instructed to respond affirmatively only when their workload would allow it. When the ready prompt was accepted an SA question was presented on the touch screen. The participant selected the answer by touching one of the alternatives shown. If the ready prompt was not accepted after one minute, it was withdrawn and presented again after two minutes, thus preserving the 3- minute interval.

Probe Question Development

For each scenario, 12-16 probe questions were presented based on scenario length. Four of the probe questions asked for workload ratings, leaving 8-12 SA questions per scenario. The probe question categories used in each semester were conflicts and sector status. Conflicts ask about existing or potential conflicts, including information relative to conflict detection. Status queries asked about current traffic, equipage mixtures, commands and communications, etc. (see examples shown in Appendix A). Probe questions were developed as follows (for more detail, see Strybel et al.,2011): First, we developed counterbalancing schemes so that all probe categories were presented equally often during the scenarios. Then we developed probe "stem" questions, in consultation with subject matter experts. These questions were general, and could be asked at any time in the scenario, or specific, requiring additional information such as aircraft callsigns based on time in the scenario. The wording of questions was vetted with subject matter experts. Once stem questions were finalized, the specific question was inserted into the appropriate time slot and additional detail (i.e., aircraft call sign or waypoint name) was added if necessary. Probe questions were scored by comparing the answers to the questions with a recording of the airspace at the time of the question. All questions were scored by at least two researchers independently. If disagreement was found a third researcher independently reviewed the question and reconciled the answer.

Results

Data Screening and Formatting

Only RTs from correctly answered probe questions (71.3% of all probes) were included for further analysis. Additionally, any probe questions with RTs greater than 40 seconds were considered outliers and exclude from the analysis. The remaining questions were placed in one of nine categories (see Appendix A). Categories were created a priori based on the existing probe question data to best represent general aspects of performance in the ATC simulations. Several such sets of categories were created, each containing between 8-12 categories. These category sets were each subjected to PCA and the set shown in Table 1 was determined to have the most variance explained with the fewest number of components as well as the clearest loadings on the resulting components.

Factor Analysis

A Pearson's correlation including all categories found that all 9 categories correlated at least .3 with at least one other item, and 6 categories had correlations of at least .4 with 1 other item, suggesting moderately reasonable factorability. Since some interns did not receive questions within a particular category, or incorrectly answered probe questions within a category, 32 interns were excluded by the listwise comparison that was used because they had missing data for at least 1 category. This left 39 participants in the analysis. Additionally, the Kaiser-Meyer-Olkin measure of sampling adequacy was .61, which is above the commonly accepted cutoff point of .6. Bartlett's test of sphericity was also significant ($\chi^2(36) = 74.98$, $p < .05$). Finally, the communalities were all above .4, further confirming that each item shared some common variance with other items. Given these overall indicators, factor analysis was conducted with all 9 items. A principle-components factor analysis using Varimax rotation and listwise comparisons was conducted, with three factors with Eigenvalues above 1 (2.64, 1.63, and 1.36, respectively) explaining 63% of the variance.

The component loading matrix for this final solution is presented in Table 1. We designated the component with loadings from probe questions related to AC separation, Global AC traffic conditions, General amount of upcoming ATC actions, and Next ATC action as "ATC Relevant" information. A second component has clear loadings with Lateral AC Distances, Altitude Distances, and Future Lateral Distances and was therefore designated "Distance Relations" information. A third component had strong loadings on probe questions related to AC Entering the Sector and Memory for Prior Conflicts. Both categories refer to probe questions unrelated to events in the sector at the time of the question, the component was therefore designated "Low Priority" information.

Composite scores were created for each of the three factors by calculating the mean for all questions from categories which had their primary loadings on each factor. Descriptive statistics are presented in Table 2. Response Times (RTs) were longest for Distance Relations probe questions followed by Action Relevant questions. Low Priority probe questions had the fastest RTs – this was likely due to the simplicity of these probe questions, which did not required a comparison or distance judgment of any type. As expected for RTs, data was positively skewed in each component, but the amount of skew ranged from very good to acceptable levels. Kurtosis was also in the acceptable range for all components.

Table 1.
*Component loadings, proportion of variance, and communalities based on PCA with Varimax rotation for 9 items from probes (*N *= 39)*

	Action Relevant	Distance Relations	Low Priority	Communality
Closest AC separation	.79			.66
Global AC Traffic conditions	.77			.70
General amount of upcoming ATC actions	.65			.52
Next ATC action	.55			.45
Lateral distance from AC to waypoint/AC		.82		.73
Altitude distances between AC		.82		.75
Future lateral distances from AC to waypoint/AC		.57		.47
AC Entering the sector			.82	.70
Memory for prior conflicts			.80	.66
Proportion of Variance	.243	.221	.161	.625

Note. Factor loadings < .4 are suppressed

Table 2.
*Descriptive statistics of responses times (RTs) in seconds to probe questions in each category component (*N *= 39)*

	Num of Items	*M* (*SD*)	Skewness	Kurtosis
Action Relevant	4	12 (3.1)	1.08	1.8
Distance Relations	3	14 (3.2)	.66	.91
Low priority	2	9 (3.3)	.66	-.22

Overall, these analyses indicated that three distinct components were underlying probe questions RTs. An approximately normal distribution was evident for the composite score data in the current study. The data were therefore well-suited for parametric statistical analyses.

Discussion

The current analyses provide a data-driven scheme for categorizing probe questions related to SA that can be used in future evaluations of NextGen concepts and technologies. The components of situational awareness generated by the PCA, Action Relevant, Distance Relations, and Low Priority bear some resemblance to existing models of SA. In particular, Probe questions loading on Action Relevant component were related to events such as existing or possible conflicts that must be kept in a state of action-readiness. The Distance Relation component loaded with probe questions that were specific to judgments of distances between items within the sector that ATCs were watching, but these items were not actionable and may not be actively stored in the mind but rather, referenced from the display when needed. These questions would take longer to answer because it would require a visual search, and indeed Distance Relation questions had the highest RTs. Together, Action Relevant and Distance Relation components are consistent with a situated SA approach in which information is represented in the head and in the operator's task environment respectively (Salmon et al., 2010; Chiappe et al., 2012). Lastly, Low Priority probe questions involve information that is either spatially or temporally outside of the current sector.

Although the number of participants in the analysis was relatively low and data was collapsed across several significant manipulations such as equipage levels and training period, we were able to extract clear components related to situational awareness. The next step will be to use this PCA and more formal data modeling to further refine SA probe question techniques. For example, it may be a more effective use of time to exclude Low Priority questions when gauging SA because these questions explain the smallest proportion of variance and do not provide a clear theoretical component of SA. Additionally, the components will be used to predict performance within the ATC simulations, including conflict resolutions and losses of separation (LOS). Ultimately, a data-driven model of SA will further resolve discrepancies between theory-based models and accommodate additional factors such as workload and ATC expertise.

References

Bacon. L.P., Strybel, T.Z., Vu, K-P., L., Kraut, J. M., Nguyen, J., Battiste, V. and Johnson, W. (2011). Situation awareness, workload, and performance in midterm NEXTGEN: effect of variations in aircraft equipage levels between scenarios. *Proceedings of the 16th International Symposium on Aviation Psychology*, Dayton, OH. 32-37, CD-ROM.

Chiappe, D.L., Strybel, T.Z., Vu, K.-P.L. (2012). Mechanisms for the acquisition of situation awareness in situated agents. *Theoretical Issues in Ergonomic Science, 13, 625-647*.

Dao A. Q., Brandt, S., Battiste, V., Vu, K-P.L., Strybel, T.Z. & Johnson, W.W. (2009), The impact of automation assisted aircraft separation on situation awareness. In M.J. Smith and G. Salvendy (Eds.): *Human Interface, Part II*, HCII 2009 Lecture Notes in Computer Science 5618, 738–747.

D'Arcy, J-F. & DellaRocco, P.S. (2001). *Air Traffic Control Specialist Decision Making and Strategic Planning – A Field Survey*. New Jersey: Federal Aviation Administration.

Durso, F.T. & Dattel, A.R. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 137–154). Aldershot, UK: Ashgate.

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

Endsley, M.R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors,* 37, 65-84.

Endsley, M.R., Sollenberger, R., & Stein, E. (2000). Situation awareness: A comparison of measures. *Proceedings of the Human Performance, Situation Awareness and Automation: User Centered Design for the New Millennium Conference*.

Gronlund, S.D., Ohrt, D., Dougherty, M., Perry, J.L. & Manning C.A. (1998). *Aircraft Importance and its Relevance to Situation Awareness.* Report DOT/FAA/AM-98/16. Federal Aviation Administration.

Salmon, P.M., Stanton, N.A., Walker, G.H. & Jenkins,D.P. (2009). *Distributed Situation Awareness: Theory, Measurement and Application to Teamwork*. Surrey, England:Ashgate Publishing.

Strybel, T. Z., Minakata, K., Nguyen, J., Pierce, R. & Vu, K-P., L., (2009), Optimizing online situation awareness probes in air traffic management tasks, In M. J. Smith and G. Salvendy (Eds.): Human Interface, Part II,

HCII 2009, *Lecture Notes in Computer Science*, 5618, 845-854.
Strybel, T.Z., Vu, K-P.L., Chiappe, D., Battiste, V., Johnson, W. & Dao, A-Q. (2011). *Metrics for situation awareness, workload, and performance: manual for online probe administration*. NASA/TM—2011–215991. National Aeronautics and Space Administration.
Strybel, T.Z., Vu, K.P. L., Battiste, V. & Johnson, W. (2013). Measuring the impact of NextGen operating concepts for separation assurance on pilot situation awareness and workload. *International Journal of Aviation Psychology, 23, 1-26.*

Acknowledgements

This project was supported by NASA cooperative agreement NNX09AU66A, Group 5 University Research Center: Center for Human Factors in Advanced Aeronautics Technologies (Brenda Collins, Technical Monitor).

Appendix A.

Sample probe stem questions with categories based on PCA and original task categories.

Sample Probe Stem Questions	Categories	Component
Estimate the lateral separation between the two closest co-altitude aircraft	Closest AC separation	Action Relevant
In the next 3 minutes (20 miles) will the majority of AC be coming from the West?	Global AC Traffic conditions	
In what general direction are the majority of overflights headed at this moment?		
How many conflicts will you resolve in the next 3 minutes (20 miles)?	General amount of upcoming ATC actions	
Will you solve any conflicts in the next 3 minutes (20 miles)?		
In what area will the next conflict occur if no further action is taken?	Next ATC action	
If AAL1320 is 2,000ft below its current altitude, how many AC will be in conflict with it in the next 3 minutes (20 miles)?		
How many miles before [callsign] reaches [waypoint]?	Lateral distance from AC to waypoint	Distance Relations
How many miles is [callsign] from [waypoint]?		
Is [callsign] higher in altitude than [callsign]?	Altitude distances between AC	
Is [callsign] lower in altitude than [callsign]?		
Will [callsign] be the next to cross [waypoint]?	Future lateral distances from AC to waypoint/AC	
Will [callsign] and [callsign] have less than 10 miles of separation if no further action is taken?		
From which direction will the next aircraft enter your sector?	AC Entering the sector	Low Priority
In the next 3 minutes (20 miles), will the majority of AC be entering the sector from the West?		
Have you moved any AC for traffic in the last 3 minutes (30 miles)?	Memory for prior conflicts	
How many conflicts have you resolved so far?		