

Situation Awareness and Performance of Student versus Experienced Air Traffic Controllers

Kim-Phuong L. Vu¹, Katsumi Minakata¹, Jimmy Nguyen¹, Josh Kraut¹,
Hamzah Raza¹, Vernol Battiste², and Thomas Z. Strybel¹

¹ California State University Long Beach,
Center for the Study of Advanced Aeronautic Technologies
1250 N Bellflower Blvd. Long Beach, CA 90840, USA

² San Jose State University Foundation and NASA Ames Research Center
Moffett Field, CA 94035, United States of America
{kvu8, tstrybel}@csulb.edu,
{kminakata, mrjimnguyen, krautjosh, hraza84}@gmail.com,
vern.battiste-1@nasa.gov

Abstract. A human-in-the-loop simulation was conducted to examine performance, workload, and situation awareness of students and retired air traffic controllers using an on-line situation awareness probe technique. Performance of the students did not differ from the controllers on many of the performance variables examined, a finding attributed to extensive sector-specific simulation training provided to the students. Both students and controllers indicated that workload was higher and situation awareness was lower in scenarios where the traffic density was high. However, the subjective workload and situation awareness scores indicate that students were more negatively affected by traffic density. Implications of these findings are discussed.

Keywords: situation awareness, air traffic controllers, NextGen.

1 Introduction

The Next Generation Airspace Transportation System (NextGen) is a transformation of the existing national airspace system in the US brought about by unprecedented growth in the demand for air travel that will ultimately exceed current day capabilities [1]. The NextGen transformations will include tools and automation that impact the roles and responsibilities of air traffic controllers (ATCs) and pilots. For example, ATCs are likely to be equipped with automation tools that enable them to safely and effectively share responsibility for separation assurance with aircrews and/or automated separation assurance systems. Presently, NextGen concepts of operations and technologies are still in development. These concepts and technologies will need to be evaluated to assess their impact on operator workload and situation awareness. Mental workload refers to the task demands placed on the human operator. Although there is still disagreement regarding the construct of mental workload, mental workload measurement (e.g., NASA-TLX [2]) has been well established [3].

Intuitively, operators know what is meant by situation awareness: controllers refer to it as “having the picture;” and pilots have called it “staying ahead of the aircraft” [4]. Although situation awareness is an accepted construct, its definition has not been agreed upon. Endsley [5] defines situation awareness as “The perception of 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” (p. 36). This definition focuses on an operator’s mental processes rather than emphasizing the end state of “awareness” [6]. Durso et al. [7] indicated that emphasis should be placed on comprehension rather than just awareness because comprehension allows for the influence of both explicit and implicit knowledge. Moreover, other authors favor definitions that place more emphasis on the dynamic aspects of situation awareness [8].

This paper describes performance differences between students and experienced controllers in a simulation designed to examine the value of different situation awareness probe questions. The simulation used current-day radar displays, tools, and operations. Situation awareness probes were developed that varied in terms of time frame (immediate vs. future) and processing category (recall, comprehension, or subjective ratings) to uncover the dimensions of situation awareness relevant to the task. Analysis of the situation awareness probes are presented in Strybel et al.’s [9] paper. The present paper focuses on differences between students and experienced controllers in terms of workload and situation awareness. In air traffic management (ATM) research, active controllers are difficult to obtain. Therefore, many research participants are retired ATCs. If students can perform well on ATM tasks with some training, then they can be included in the pool of research participants for evaluating NextGen concepts and technologies.

2 Method

2.1 Participants

Seven students who are training for careers in ATM at Mount San Antonio College (Mt SAC) and nine retired ATCs (6 TRACON and 3 ARTCC) participated in the simulation; see Table 1 for demographic information. The students completed a course in the Air Traffic Control Environment at Mt SAC, which includes topics of aircraft characteristics, air traffic procedures, and phraseology. In addition, they completed a 16-week ATC RADAR simulation course offered in the Center for the Study of Aeronautic Technologies at California State University Long Beach. Radar simulation training was provided with the Multi Aircraft Control System (MACS) software developed by the Airspace Operations Laboratory at NASA Ames Research Center [10]. This course met for 6 hours once a week. Students participated in simulation exercises that focused on ZID Sector 91, shown in Fig. 1. Students were trained in a radar environment to accomplish the tasks of descending arrivals and climbing departures while maintaining an efficient flow of en route traffic. They were given training relating to conflict recognition and resolution, conflict avoidance via structurally placing aircraft on segregated routes, understanding the difference between habitual traffic flows and actively assessing pairs of aircraft for conflicts. Thus, the students had extensive training in the simulation environment used in the present study.

Of the experienced ATCs, TRACON controllers had an average of 19.5 years of line experience in SoCal TRACON, but had no previous experience with either ZID 91 or MACS. ARTCC ATCs had an average of 20 years of line experience, and had participated in previous experiments at NASA Ames using MACS and managing traffic in ZID 91. However, in those experiments, the controllers used advanced conflict detection and resolution tools that were not enabled in the present simulation. In sum, students had no line experience in ATM, but they had been trained extensively on MACS and ZID 91. Experienced ATCs had extensive line experience, but little experience with MACs and ZID 91 using current day tools.

Table 1. Demographics for students and experienced ATCs. Experience was based on self report, using a 1-7 Likert scale: 1= no experience to 7 = very experienced.

ITEM	EXPERTISE GROUP	
	Students	ATCs
Experience with ZID	4.57	2.11
Experience with ZKC	1.14	1.22
Experience with MACS software	3.86	2.22
Experience with radar simulation	3.33	3.00
Years of Military ATCo Exp	0	4
Years of Civilian ATCo Exp	0	23

2.2 Design

The simulation employed a 2 (Group: students vs. experienced ATCs) X 2 (Traffic Density: low vs. high) mixed factorial design. Group was a between-subjects variable. We collapsed the TRACON and en-route controllers into a single group for two reasons. First, the number of en-route controllers was low (N = 3) and second, there were little differences in performance between the experienced controllers on many of the variables examined. The dependent measures included performance, subjective workload, and accuracy and latency to situation awareness probes. Performance was assessed with the following sector performance variables: mean handoff time, standard deviation of handoff time, mean time per aircraft through sector, standard deviation of time through sector, number of aircraft through sector, and number of losses of separation. For all time-based measures (e.g., mean handoff time) inverse transformations were used to ensure normal distributions.

2.3 Apparatus

Simulation environment. The entire simulation was run using the Multi Aircraft Control System (MACS). The MACS software is a medium fidelity simulation computer application that has the ability to simulate both ground- and air-side operations [10]. Two parallel simulation worlds were created for the ATCs and each world contained eight computers running the necessary simulation components. Each ATC station had a simulated RADAR screen of sector ZID 91 that mimicked current day ATC operations. The display was augmented with a probe window that was used to present situation awareness questions. A voice server station provided a voiceIP

communication system for the controller – pilot communications. All voice communications were recorded with Creative Media Player, and were later transcribed. All aircraft in the simulation were piloted by experimental confederates who initiated and responded to ATC transmissions.

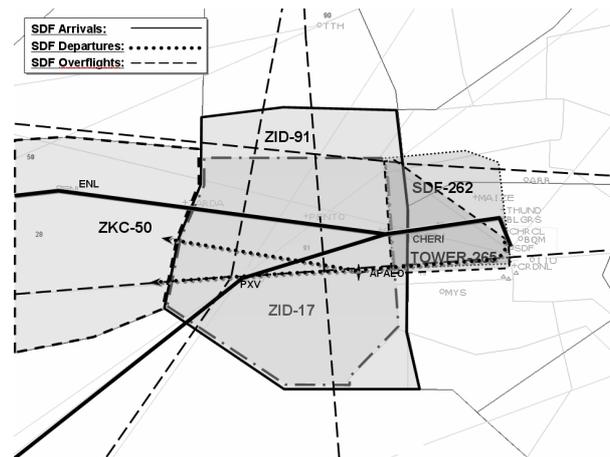


Fig. 1. Illustration of the sectors and traffic flows used in the simulation

Scenario Development. Six different scenarios were created, three of which corresponded to the low traffic (50% of current day, 1x traffic density) and three to the high traffic (75% of current day, 1x traffic density) manipulation.

SA Probe Question Development and Implementation Technique. Three information processing categories were created to reflect different components of situation awareness: subjective, recall, and comprehension. Subjective questions asked operators to rate the information being queried based on their own assessment of the situation. Recall questions are those in which the answers can be based on information in memory or looked up on the display if the operator was aware of where to look for the information. Comprehension questions were used to assess the operator's understanding of the situation, and usually required the controller to derive the answer rather than recalling the item or looking it up on the display. Within categories, the questions were divided into two time frames, those reflecting information in the immediate past or past versus questions that required projection into the future, see Strybel et al. [9] for data regarding probe categories. All probe questions required closed-ended responses, being yes/no questions, questions requiring a numeric response (0-4 and 5+), or rating questions based on a Likert scale.

Probe questions were administered using the SPAM technique developed by Durso and colleagues [11]; see Fig. 2 for probe administration and response sequence. First a "ready" prompt was presented, with the controller responding to this when his or her workload was low enough for them to accept a question. When the ready response was received, the question was presented in the probe window and controllers responded to the questions using pre-assigned response buttons on a configurable

response panel (see Fig. 2). Probe questions were sent approximately every three minutes beginning four minutes into the scenario.

Three configuration keys on the upper-right response panel were for experimenter use only. In the upper-left corner, there was a “READY” for probe question key. The remaining keys were for the possible responses that could be made to the questions, including one DK key for “Don’t Know.” The six keys on the bottom row of the panel allowed participants to respond to questions with a response of No or Yes, Likert-scale responses (e.g., Very Unlikely to Very Likely) and the numbers of 0 through 5+.

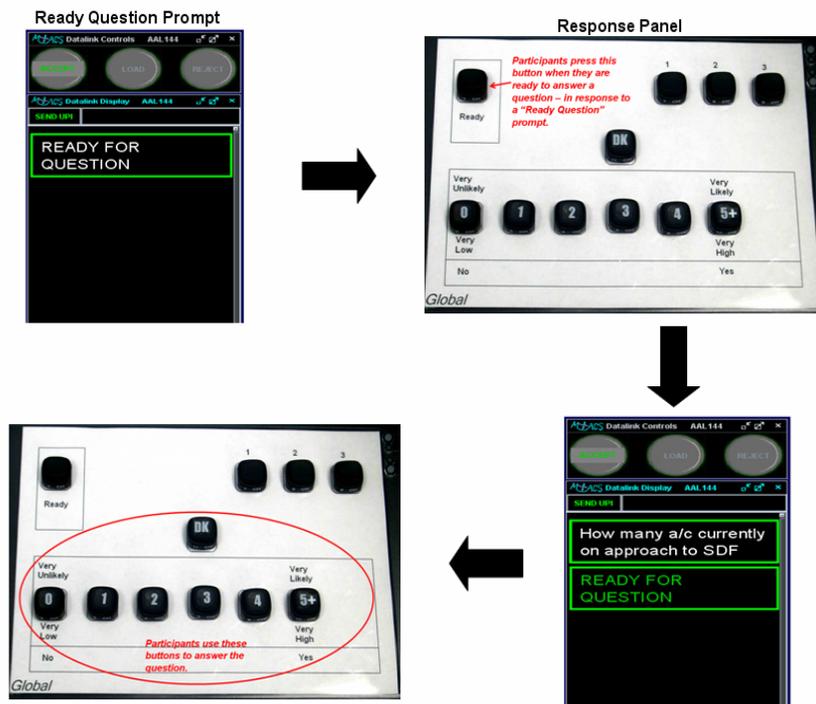


Fig. 2. Illustration of probe question administration sequence and the display and control interfaces used for responding to probe questions

SART and NASA TLX. The Situation Awareness Rating Technique (SART) [12] was utilized to capture participants’ subjective SA experiences. The SART is a subjective measure which consists of nine scales which are categorized into three subscales of Understanding, Demand, and Supply. All of the SART scales function as seven-point scales, where 1 = “Low” and 7 = “High.” A combined SART score was used as an estimate of overall SA: SART-Combined = Mean Understanding Rating – (Mean Demand Rating – Mean Supply Rating) [12].

The NASA-Task Load Index (TLX) [2] was used to collect subjective assessments of workload. The TLX consists of six subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. All dimensions of the TLX had a 15-point scale, where 0 = “Low” and 15 = “High.”

2.4 Procedure

Training. All participants were trained on the first day of the simulation with 2 hours of classroom training and 3-4 hours of hands-on simulation training. In-class training consisted of a briefing on current day ATC operations and traffic flows in sector ZID 91. It also included information regarding how to interact with the MACS software, and the roles and responsibilities of the controllers. At the end of the training, participants were trained on the probe administration technique. Hands-on simulation training was conducted with two training scenarios (at least two replicates of each), until each controller was comfortable with the ATC task and the probe procedure.

Experimental Trials. Day two consisted of the six experimental trials. During each 40-minute scenario, 12 probe questions were presented via the probe panel in about 3 minute intervals. When a scenario was complete, participants were given the NASA-TLX and SART. A 15-30 minute break was given after each scenario as well as a 1-hour lunch break. After completing all scenarios, participants filled out a post simulation questionnaire and were debriefed.

3 Results

3.1 Performance

The participants engaged in their typical traffic management behavior and some differences in sector performance were observed between students and ATCs. All measures were submitted to separate 2 (Group: students vs. controllers) x 2 (Traffic Density: low vs. high) ANOVAs, with Group as a between-subjects variable.

For mean handoff time, the main effects of group and traffic density were not significant, $F_s < 1.0$, but the interaction of the two variables was, $F(1,15) = 7.99$, $p < .013$, see Fig. 3. Students showed no difference in handoff time between low and high traffic scenarios, but ATCs took longer to handoff aircraft for high traffic level scenarios. This finding may be due to the fact that ATCs tended to be more cautious about when to handoff aircraft when traffic levels are high.

In terms of the mean time per aircraft in the sector and time in sector variability, only a marginal effect of group was obtained for the latter measure. The variability in time for an aircraft to move through the sector tended to be longer ($M = 203$ s) for students than for ATCs ($M = 186$ s), $F(1,15) = 3.90$, $p = .06$. No other variables, including losses of separation, yielded group differences. The lack of group differences is likely due to the students receiving extensive sector-specific training prior to the simulation.

In addition to examining sector performance variables, we examined voice transcriptions to determine whether there were group differences between operator behaviors on the number of commands given (separately for altitude, heading, and speed), number of queries made to the flight deck, number of corrections to pilot read-back errors, etc. However, none of these variables yielded significant group differences.

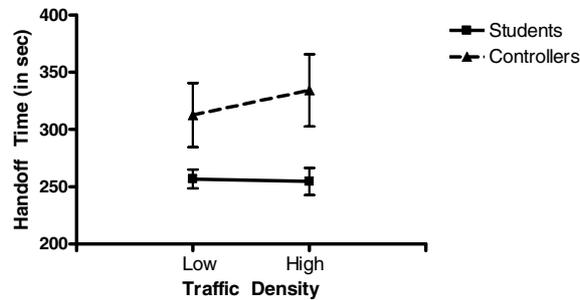


Fig. 3. Handoff Time as a Function of Traffic Density and Group

3.2 Situation Awareness

Answers to SA Probes. Students and ATCs responded differently to the probe questions. Students answered a higher number of probes ($M = 7.7$ questions per scenario) than ATCs ($M = 6.5$ questions per scenario), $t(94) = 3.28, p < .001$. ATCs also ignored more “ready” prompts ($M = 8\%$ of probe questions per scenario) relative to students ($M = 3\%$), $t(94) = 2.32, p < .05$. Furthermore, students left fewer unanswered questions after accepting the “ready” prompts ($M = 0.4\%$ of probes per scenario) compared to ATCs ($M = 1.7\%$), $t(94) = 1.80, p = .07$. Thus, students seemed to be more compliant than ATCs in answering probe questions.

Probe accuracy scores (proportion correct of answered probes) were submitted to a 2 (Group: students vs. ATCs) X 3 (Question category: recall vs. comprehension vs. subjective) X 2 (Question tense: immediate vs. future) X 2 (Traffic Density: low vs. high) mixed ANOVA. Accuracy to subjective questions was derived by comparing the participant’s answer to a standard derived by a retired ATC who has been working in the lab. Because we are primarily interested in ATC versus student performance, only the main effect of Group and its interaction with other factors will be reported. Although we report the original degrees of freedom, p -values reflect the Huynh-Feldt correction for violations of sphericity where appropriate. Moreover, all *post hoc* analyses were performed with a Bonferroni correction for multiple comparisons.

A marginal main effect of group was obtained, $F(1,14) = 3.86, p = .06$. On average, students ($M = 73\%$) were more accurate than ATCs ($M = 66\%$) on the probe questions they answered. However, this effect was moderated by a marginally significant three-way interaction of Group, Question category and Question tense $F(2,28) = 2.57, p = .09$, as shown in Fig. 4. Both students and ATCs were more accurate for recall questions directed at future events than present events. Both groups showed less agreement with our standard on subjective assessments of future than present events. For comprehension questions, students were less accurate for future events compared with present events, but ATCs were more accurate for future events than present events. However, students were more accurate than ATCs for comprehension questions. Students may have been more accurate than ATCs overall because they tried harder to answer the probes (e.g., accepted more questions, had fewer time outs, and left fewer questions unanswered).

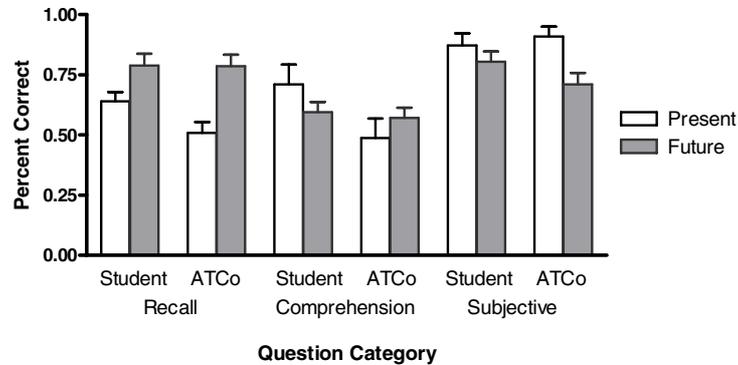


Fig. 4. Accuracy of probe questions that were answered as a function of group, question category, and question tense

Latencies to SA Probes. Latencies to correct answers on probe questions was submitted to a similar mixed ANOVA as probe accuracy. However, there was no significant effect of Group, and Group did not interact with any other factors.

Subjective Situation Awareness. SART composite scores were analyzed with a 2 (Group: students vs. ATCs) X 2 (Traffic Density: low vs. high) mixed ANOVA. The SART composite score yielded a significant traffic density x group interaction, $F(1, 14) = 4.60$, $p < .05$. Specifically, students reported having more SA when traffic density was low ($M = 7.0$) than when it was high ($M = 5.2$), $p < .01$. Similarly, ATCs reported having more SA when traffic density was low ($M = 6.9$) compared to when it was high ($M = 6.0$), $p < .05$. The main difference was that the students were more affected by scenario difficulty than were the ATCs.

3.3 Workload

NASA-TLX. Six mixed ANOVAs on each scale of the TLX were conducted. For all analyses, only the main effects of traffic density were statistically significant, $F_s(1, 14) > 27.58$, $p_s < .001$. The high density scenarios were rated higher in workload than the low density scenarios, see Table 2.

Table 2. Mean TLX workload ratings for the low and high traffic density conditions

Workload Dimension	Low Density	High Density
Mental Demand	7.70	12.33
Physical Demand	4.35	8.24
Temporal Demand	6.14	10.66
Performance	4.51	8.50
Effort	7.82	12.15
Frustration	3.78	7.88

4 Discussion

Highly trained students did not differ much from ATCs on the sector performance variables measured in this simulation. This finding is likely due to the intense, sector-specific simulation training given to the students rather than overall air traffic control abilities between the two groups. Although both students and ATCs indicated that workload was higher and situation awareness was lower in the hard than easy scenarios, students reported being more negatively effected by scenario difficulty.

In terms of situation awareness as measured by probe question accuracy, students were more accurate overall than ATCs. There are two possible reasons for this finding. The first is that sector-specific knowledge is important for situation awareness. Because students had more training with traffic flows and sector-specific characteristic, they were able to maintain more awareness of the information in the sector, or had more knowledge about where to obtain this information. Second, the students showed more compliance in answering probe questions compared to ATCs, which could make them more motivated to answer the questions correctly than ATCs. ATCs were more willing to abandon the answering of questions compared to students. Since the scenario was never frozen during probe administration, there was a possibility that critical events could appear after the controller accepts the “ready” prompt, but before the controller answered the questions. In those cases, ATCs gave more priority to the air traffic management task than to answering the probe questions, which led to more questions being abandoned compared to students. For comprehension questions, students were less accurate on questions about future events compared to present events, but ATCs showed higher accuracy for future than present events. These findings are consistent with the observation that good controllers are able to anticipate future events [13].

In general, the present simulation showed that sector-specific knowledge is very important at least in some measures of performance and situation awareness. Thus, researchers should make sure that participants are adequately trained on the specific roles and responsibilities being evaluated, and that critical aspects of experimental procedures that are not typical standard controller tasks be emphasized. In ATM research, it is difficult to recruit current FAA employees as participants. This study shows that students, with some training and experience, can perform well in ATM tasks, which allows them to be an additional source of research participants for evaluating ATM concepts in general and NextGen concepts in particular.

Acknowledgements. This simulation was partially supported by NASA cooperative agreement NNA06CN30A.

References

1. JPDO, Concept of Operations for the Next Generation Air Transportation System, V2.0 (June 2007)
2. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.), pp. 139–183. North-Holland, Amsterdam (1988)

3. Pickup, L., Wilson, J.R., Sharpies, S., Norris, B., et al.: Fundamental examination of mental workload in the rail industry. *Theor. Issues in Ergon. Sci.* 6, 463–482 (2007)
4. Jeannot, J.: Situation Awareness: Synthesis of Literature Search. EEC Note 16/00, Eurocontrol Experimentale Center (2000)
5. Endsley, M.R.: Measurement of situation awareness in dynamic systems. *Human Factors* 37(1), 65–84 (1995)
6. Banbury, S., Tremblay, S.: A cognitive approach to situation awareness: Theory and application. Ashgate, Farnham (2004)
7. Durso, F.T., Rawson, K.A., Girotto, S.: Comprehension and situation awareness. In: *Handbook of applied cognition*, 2nd edn., FT Durso, pp. 163–193. Wiley, Hoboken NJ (2007)
8. Salmon, P.M., Stanton, N.A., Walker, G.H., Baber, C., Jenkins, D.P., McMaster, R., Young, M.S.: What really is going on? Review of situation awareness models for individuals and teams. *Theor. Issues in Ergon. Sci.* 9, 297–323 (2008)
9. Strybel, T.S., Minakata, K., Nguyen, J., Pierce, R., Vu, K.-P.L.: Optimizing online situation awareness probes in air traffic management tasks. In: Smith, M.J., Salvendy, G. (eds.) *Human Interface, Part II, HCII 2009*. LNCS, vol. 5618, pp. 865–874. Springer, Heidelberg (2009)
10. Prevot, T.: Exploring the many perspectives of distributed air traffic management: The multi aircraft control system MACS. In: *International Conf. on Human-Computer Interaction in Aeronautics, HCI-Aero 2002*, October 23–25. MIT, Cambridge (2002)
11. Durso, F.T., Bleckley, M.K., Dattel, A.R.: Does SA add to the validity of cognitive tests? *Hum. Factors* 48, 721–733 (2006)
12. Taylor, R.M.: Situational awareness rating technique (SART): The development of a tool for aircrew systems design. *Situational Awareness in Aerospace Operations*, AGARD-CP-478 (1990)
13. D’Arcy, J.-F., Della Rocco, P.S.: *Air Traffic Control Specialist Decision Making and Strategic Planning*. National Technical Information Service, Spring Field (2001)