MEASURING THE IMPACT OF RADAR SUPPORT OPTIONS ON AIR TRAFFIC CONTROL SITUATIONAL AWARENESS

Dr John Steele
Op Research Analyst/DGLCD*
Fort Frontenac, PO Box 17000 Station Forces
Kingston, Ont. Canada K7K 7B4
(613) 541-5010 x 8717
Steele.JA2@forces.gc.ca

Mr Roger L. Roy (POC)
Senior Op Research Analyst/DGLCD
Fort Frontenac, PO Box 17000 Station Forces
Kingston, Ont. Canada K7K 7B4
(613) 541-5010 x 8725
Roy.RL@forces.gc.ca

Ms Eugenia Kalantzis
Op Research Analyst/DGLCD
Fort Frontenac, PO Box 17000 Station Forces
Kingston, Ont. Canada K7K 7B4
(613) 541-5010 x 2469
Kalantzis.E@forces.gc.ca

Maj James S. Denford
Director Land Synthetic Environments 4/LFDTS*
PO Box 17000 Station Forces
Kingston, Ont. Canada K7K 7B4
(613) 541-5010 x 8681
Denford.JS@forces.gc.ca

* DGLCD=Director General Land Combat Development
* LFDTS=Land Forces Doctrine and Training System

Abstract

Airspace restrictions in Kabul in 2003 limited the potential capability of the Tactical Unmanned Air Vehicle (TUAV) within the area of operations (AO) of the Kabul Multinational Brigade (KMNB). A Limited Objective Experiment was conducted to examine five different radar support options and three different information display options to increase the airspace situational awareness of the air traffic control officer (ATCO). The use of techniques such as SAGAT, SART and NASA-TLX were operationalized and proved to be effective in determining differences in situational awareness (SA) and situational understanding (SU). These three measures appear to have great utility in a number of domains and should be investigated further. Rapid report generation was made possible through advances in simultaneous data capture, and the automation of the data consolidation and analysis routines. The Mann-Whitney U test for statistical significance was found to be useful because of the nature and limited size of the data sets collected. The employment of any radar in support of air traffic control (ATC) operations significantly improved the Level 1 SA of the ATCO. Only the MPN-25 radar option showed statistically significant improvements in Level 2 and Level 3 Tower SA. The MPN-25 SART SU scores and NASA-TLX workload scores were significantly better than all other radar support options. The decrease in workload associated with the provision of an SA aid (display) allowed the ATCO to better leverage his SA into SU. Based on these findings, and considering the limitations of the simulation, the MPN-25 provided the best radar support option, but any of the radars examined in this experiment would assist in the establishment of positive control over TUAV operations in the Kabul International Airport (KIA) controlled airspace.

INTRODUCTION

Air traffic control (ATC) specialists are called upon to identify and project the paths of an ever-fluctuating number of aircraft in order to ensure goals of minimum aircraft separation, as well as safe and efficient take-off, en route and landing operations. The success of the air traffic control officer (ATCO) in this task depends upon his/her level of situational awareness (SA). These controllers must maintain an up-to-date assessment of the rapidly changing location of each aircraft (in three-dimensions), its projected location, the relative positions of aircraft to one another, along with other relevant aircraft parameters such as destination, speed, fuel, altitude, and heading. The central skill of the controller is the ability to respond to a variety of quantitative inputs about several aircraft simultaneously and to manage a continuously changing mental picture that forms the basis for planning and controlling the course of the aircraft. Providing controllers with a precise, complete, and up-to-date picture of the situation may prove to be a daunting challenge, as the environment in which they work becomes even more complex and demanding (Endsley and Rodgers, 1994).

The Minister of National Defence approved the Urgent Operational Requirement to procure a Tactical Unmanned Air Vehicle (TUAV) on 22 May 03. The acquired Sperwer TUAV system included four Air Vehicles (AV), two Ground Control Stations (GCS) with Ground Data Terminals, one launch system, a Platoon Maintenance Facility and ancillary equipment. Prior to deployment, the TUAVs were restricted in their operations in the vicinity of the airfield. As depicted in Figure 1, a no-fly zone was established within the area directly surrounding the airfield, 8 km beyond either end of the runway and 4 km to either side of the runway; TUAV flight was restricted to a height of 500ft in the area extending 16 km beyond either end of the runway and 7 km to either side of the runway; TUAVs were free to fly to a height of 2500 feet above ground level (AGL) and as close as 4 nautical miles from the airport in the remaining Area of Operations (AO). However, the commander has the authority to suspend air traffic and provide the TUAV with freedom of flight when it is deemed mission critical.

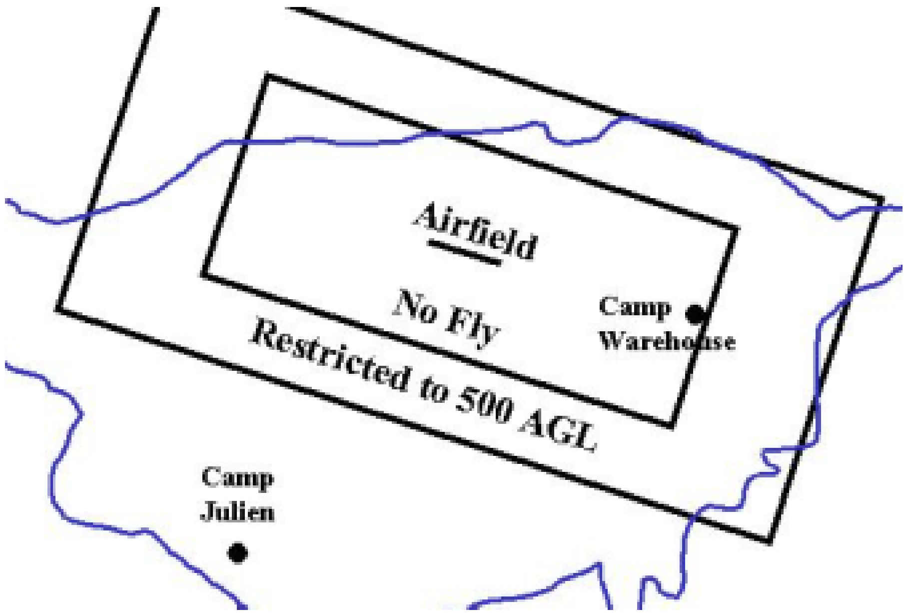

Figure 1. TUAV Airspace Restrictions Over Kabul

Three candidate sites for radar installations were identified for this experiment: the Kabul International Airport (KIA) for civilian ATC radars, and either Camp Warehouse or Camp Julien for military radar assets. The relative position of these three sites within the AO is illustrated in Figure 2, which displays a hillshade view of the area. The selection of both Camp Julien and Camp Warehouse as potential radar sites is fuelled partly by the presence of a mountain range running through the centre of the AO, severely limiting the visibility of an aircraft over the entire AO from either one of these two locations.

Six radar options were investigated: 1) Baseline (no radar); 2) Quad ATC Radar at KIA; 3) MPN-25 ATC Radar at KIA; 4) Air Defence/Anti Tank Systems (ADATS) at Camps Julien and Warehouse; 5) Skyguard Fire Control Units (FCUs) at Camps Julien and Warehouse; and 6) Quad ATC Radar at KIA and Skyguard FCU at Camp Julien. Three different types of information displays within the ATC tower were studied: a TUAV moving map display, the MPN-25 radar situation display (RSD) and the Air Defence System Integrator (ADSI) display.

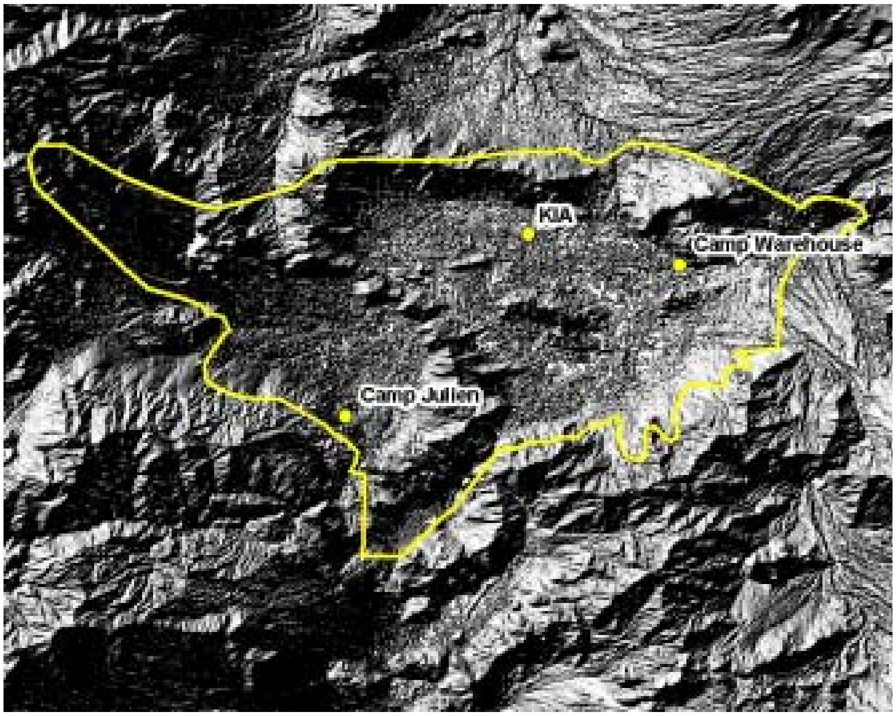

Figure 2. Hillshade View of the KMNB Area of Operations

THEORY

While several definitions of situational awareness (SA) have been offered, the most generally applicable definition is that provided by Endsley (1988). SA is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". SA can be both a product and a process. The product is the ultimate state of awareness the individual has of the system and its environment; the process is the combination of perceptual and cognitive tasks required to maintain the product.

The primary method for measuring SA during the experiment was the Situational Awareness Global Assessment Technique (SAGAT) (Endsley, 1987). When using SAGAT, the simulation is frozen at random intervals and data is collected on three levels of SA. Endsley defines the three levels of SA (perception, comprehension and prediction) , which were evaluated in this experiment. The first level of SA, referred to as Level 1 SA, involves perceiving the current status, attributes, and dynamics of relevant elements in the environment. In this study, Level 1 SA included aircraft type, call sign, location, altitude, altitude change, airspeed, heading, heading change, intention, and emergency status. For each parameter, criteria were developed to measure the controller's response relative to reality, and to assign a score between 0 and 5 to these responses. Subject matter experts participating in the experiment provided their rating of the impact of each of these parameters on air traffic safety. The results were used to derive weight factors that were then used to combine the ten assigned scores into a single SAGAT Level 1 SA score out of 50.

Level 2 SA extends beyond the perception of current facts to include the comprehension of their significance in light of the controller's goals. Since the focus of this experiment was to determine whether radar supported SA was adequate to ensure safety when a remotely piloted vehicle was inside controlled airspace, the measure of Level 2 SA was selected as the accuracy in the knowledge of which of the aircraft were closest to the TUAV at any given time. Again, criteria were developed to assign scores out of 10 to measure the correctness of the identification of aircraft proximity.

Level 3 SA is the projection of the future actions of the elements of the environment in the near term. The corresponding measure used in this experiment was derived from the response to the Level 2 SA question in time frames of 2 and 5 minutes in the future, i.e. the correctness of the prediction of aircraft proximity to the TUAV in 2 minutes, and in 5 minutes from the present. These measures were scored in the same manner as for the Level 2 SA parameters.

A second more subjective technique used was the Situation Awareness Rating Technique (SART) (Taylor, 1989) in which the ATCO assesses ten different dimensions of the quality of his own SA on a scale from one to seven. These dimensions can be grouped and interpreted as indications of supply of attention, demands upon attention, and degree of situational understanding (SU).

Finally, a technique used to assess the changes in the ATCO workload introduced by radar support, was the NASA Task Load Index (TLX). This technique required participants to assign a Likert scale rating from 1 to 7 to each of six different dimensions of the burden involved in the task at hand, and to conduct pair-wise comparisons between the different dimensions, identifying which of the two has a greater contribution to task burden. From these ratings, a single TLX score is calculated. The dimensions rated include mental demand, physical demand, time pressure, effort required, pressure due to performance level, and level of frustration.

SYNTHETIC ENVIRONMENT

The intent of the synthetic environment (SE) was to provide sufficient contextual cues to stimulate expert performance in the experiment participants. Elements simulated included the ATC tower, the Quad or AN/MPN-25 ATC radar, Land Force ADATS or Skyguard Fire Control Unit, a Sperwer TUAV and Ground Control Station (GCS), and an exercise control (EXCON) element driving air traffic in the Kabul area.

Two pilots were employed in EXCON to direct the movement of aircraft within the OneSAF Test bed Baseline (OTB) and to simulate radio communications through two ModIOS stations. Flight profiles were developed using the Jeppesen diagrams for KIA, augmented by recommendations from two CC-130 Hercules pilots with recent experience in Afghanistan. The employment of pilots qualified in both fixed and rotary wing aircraft was vital to the establishment of a realistic virtual environment as their voice procedures helped better situate the ATCO in his ATC tasks.

The Sperwer TUAV simulation consisted of an OTB workstation that simulated the Ground Control Station (GCS) directing the flight path of the TUAV, and a ModIOS moving map display showing the TUAV position over the terrain, the view from a nose mounted camera, and the Orientable Line-of-Site Payload (OLOSP) feed for the payload operator. Radio communications were provided between the tower and the GCS. The constructive simulation permitted the TUAV controller to fly the TUAV through various legs at various altitudes and speeds, to take off and land the TUAV, and to initiate appropriate immediate actions. The air vehicle and payload characteristics were modeled to specifications of the Statement of Operational Requirement (SOR) for the system.

The tower simulation consisted of a workspace that included a full 360-degree view from the tower, represented by three ModIOS presentations across six screens. The ATCO and assistant were provided with a station to simulate the use of binoculars from the tower position. During the examination of different information display options presented to the ATCO in the tower, a ModIOS station was provided for remote display of the MPN-25 plan position indicator (PPI), or the moving map display from the TUAV. The Air Defence System Integrator (ADSI) was simulated using an OTB screen showing all flight operations and representing full coverage of the area (integrating radar coverage from Bagram, Kandahar and Kabul). ATC communications were provided via telephone links to the AD radars, and radio links to the TUAV GCS and the various aircraft.

The air defence radar simulation consisted of a workspace that included two PPIs and two electro-optical displays. Telephone communications were provided to the AD Battery Command Post and the radar operators maintained listening watch on the tower radio frequency. The radar and optics performance characteristics were modelled to specifications provided by the manufacturer.

The ATC radar simulation consisted of a workstation that included an operator with a PPI for Quad radar operations, with remote connection to the tower location for MPN-25 operations. Telephone communications were provided to the tower, and the radar operator maintained listening watch on the tower radio frequency. The radars' performance characteristics were modeled for Primary Search Radar operations to specifications provided by the ATC squadron employing the systems.

THE EXPERIMENT

Elements incorporated in the experiment were manipulated to simulate the interaction between the radar, aircraft, and the ATCO. The aircraft in the simulation included a variety of domestic and international commercial, military and private air vehicles. Boeing 707s, DC-3s and Hercules transports shared the airspace with Cessnas, Helicopters and the TUAV. Circumstances modeled included mechanical, weather and medical emergencies, as well as situations where an aircraft was unable or unwilling to communicate with the tower.

Each half-hour scenario involved a total of 12 aircraft on a variety of inbound, outbound, en route and local flight paths. Air traffic tempo was regulated to maintain a moderate load of between three and six aircraft within the ATCO's control at all times during the simulation. It was determined that only about four iterations of each radar option could be completed in the allotted time. This resulted in a very tight schedule with four iterations completed in the morning, four (sometimes five) in the afternoon and two (sometimes three) in the evening. A total of 24 iterations plus three excursions were completed.

Following each of the 24 iterations, participants submitted their responses in a shared Microsoft Excel workbook, which contained four worksheets for the ATCO (three for SAGAT and one for

SART and TLX) and three for EXCON (SAGAT only)). This methodology minimized the need for analysts to transfer data, eliminated the possibility of errors in data transfers, and provided rapid results for comparison and interpretation. The EXCEL worksheets were pre-programmed to allow for automatic data manipulation using a series of macros written in Visual Basic for Applications (VBA). Each of the six configurations were run four times with three pauses each, yielding 12 SAGAT data points per configuration.

ANALYSIS

The analysis of the data involved several stages, each described in a subsection below.

Mapping ATCO-listed to EXCON-listed aircraft

Since the queries to the ATCO were not multiple choice, some decision had to be made about which of the aircraft listed by EXCON was being described by the data for each aircraft on the ATCOs inputs. The scores assigned to SAGAT data from the ATCO will rise or fall depending upon which correspondences are assumed. To make these choices, a macro was written which tested every possible set of mappings from ATCO aircraft to EXCON aircraft. An objective function was evaluated for each set of mappings. This function was a weighted sum of scores assigned to 4 data items: 40% on each of location and aircraft type and 10% on each of altitude and speed. The mapping chosen for evaluation of SA was the one that optimized this function.

SAGAT Level 1 SA items

Criteria were developed to rate the controller's estimate of the various aircraft parameters. For each of the 10 Level 1 SA parameters, a score level between 0 and 5 was assigned. This was done during each pause in the simulation.

Call Sign. A macro was created to systematically compare the call sign listed by the ATCO with the actual call sign of the aircraft, and to penalize character transpositions and substitutions. Although this was trialed and proved to be somewhat successful, it was too computationally intensive for repetitive use and was abandoned in favour of a more rudimentary scheme using the EXCEL VLOOKUP function. The ATCO call sign response was inserted into the appropriate location in an alphabetized list of call signs, and one point is subtracted for each position that the call sign is from the true call sign.

Location. The ATCO indicated aircraft locations by recording a number for each aircraft in an EXCEL-based graphical overlay displaying the positions of the KIA runway and Camps Warehouse and Julien. Underlying this graphical overlay, the EXCEL sheet was formatted such that each worksheet cell represented a 2.5km square. Participants were penalized for every 5 km of straight-line distance between the ATCO location for an aircraft and the EXCON location for the same aircraft.

Altitude. Aircraft altitude was scored using the absolute value of altitude differences. As the difference between the EXCON and the ATCO altitude values varied between 0ft to 500ft, 1000ft, 1500ft, 2000 ft, and 3000ft or more, the altitude score fell from 5 to 4, 3, 2, 1, and 0, respectively.

Aircraft Type. The six types of aircraft in the simulation were grouped as transport (Boeing 707, DC-3 and Hercules), small (Cessna and Griffon helicopter) and TUAV. The size of scoring penalty was correlated with the seriousness of the consequences of the error in identification. Full marks were allocated only with correct aircraft identification. Cases where transports were confused with other transports were given 4 points. Confusion between two different small aircraft was given 3 points. Confusing a transport with a small aircraft, and vice versa, was given 2 points. Confusing a small aircraft with a TUAV, and vice versa, was given 1 point, and confusing a transport with a TUAV, and vice versa, was given 0 points.

Speed. Aircraft speed was scored using the absolute value of speed differences. When the difference between the speed of an aircraft given by the ATCO and EXCON exceeded 25kts, 50kts, 75kts, 100kts, and 150kts, speed scores fell to 4, 3, 2, 1, and 0, respectively.

Heading. Aircraft heading was scored using the absolute value of heading differences, while ensuring that errors were corrected around the 0/360 degree mark, i.e. 355 degrees was interpreted as 10 degrees away from 5 degrees, and not 350 degrees. As the difference between the heading of an aircraft given by the ATCO and EXCON exceeded 15 deg, 30 deg, 45 deg, 60 deg, and 90 deg, the score assigned fell to 4, 3, 2, 1, and 0, respectively.

Heading Change. A score of 5 was given for correctly identifying left or right turning or straight flight. Three points were given for confusing a turn with straight flight, or vice versa, and a score of 1 was given for confusing a left turn with a right turn, and vice versa.

Altitude Change. As with the aircraft heading change score, correctly identifying a climbing, descending or level flying aircraft earned 5 points. Confusion between level flight and either climbing or descending was given 3 points. Indicating a climbing aircraft to be descending, and vice versa, earned only 1 point.

Activity. Correctly identifying whether the aircraft was en route, inbound, outbound or local earned 5 points. Calling an inbound aircraft outbound, and vice versa, earned 3 points because it was considered that, although not correct, the participant was aware that activity revolved around the runway. Confusing local aircraft with in- or outbound aircraft, and vice versa, was deemed to indicate more serious confusion and given 2 points. Indicating that an en route aircraft was not en route resulted in 1 point, and believing an aircraft not en route to be en route was deemed most serious as interactions with other aircraft would not be anticipated and was given 0 points.

Emergency Status. Correctly identifying the presence or absence of an aircraft emergency was given 5 points. Incorrectly ascribing an emergency was considered less serious than incorrectly ascribing no emergency. The former earned 1 point and the latter 0 points.

Defining a Global SAGAT Level 1 SA Score

To simply average the Level 1 SA parameter scores out of 5 would be to falsely suggest that all ten items were of equal importance in determining Level 1 SA. Instead, subject matter experts participating in the experiment were asked to identify the level of importance of these parameters by completing the 45 possible pair-wise comparisons, indicating which of each pair of items was more important for air safety. From these pair-wise comparisons, individual priority ratings were determined between the ten items. One option was to compute an unweighted average of the row sums from all of the players (Uniform weights). A second option was to compute the weighted average of the row sums from all the players. This yielded what were called Equal Player weights. The other option considered was based on the assumption that the best subject matter expert on the importance of SA items for air safety was the ATCO. Rather than using his feedback exclusively, his responses were assigned a 50% influence and those of the other participants were assigned 12.5% influence each. Using these weight factors, the row sums were combined to give what was called a set of ATCO Dominant weights. The resulting weights were used to combine the ten Level 1 SA items to give a Global Level 1 SA score out of 50.

SAGAT Level 2 and 3 Scores

Six, three and one points were available as component scores for correct listing of the first, second and third closest aircraft, respectively. The portions of each of these components assigned are given in Table 1. Listing the three closest aircraft but in the wrong order assures at least 4 out of 10.

Table 1. *SAGAT Level 2 and 3 Scoring Scheme*

	ATCO Listing			
EXCON Listing	**Nearest**	**2nd Nearest**	**3rd Nearest**	**Not Listed**
Nearest	6	3	1	0
2nd Nearest	2	3	1	0
3rd Nearest	1	1	1	0

SART Scores

The Situational Awareness Rating Technique (SART) requires the ATCO to rate ten different aspects of his own SA. Analysis on the basis of the ten different dimensions is known as 10-D SART. The responses can then be averaged within three groups to indicate the demands on attention, the supply of attention, and the degree of situational understanding. Analysis of these three aggregate measures is called 3-D SART. The 10-D items rated are shown in Table 2.

Table 2. *The Components of the SART*

3-D	**10-D**	**Description**
Demands on Attention	Instability of Situation	Situation's likeliness to change suddenly
	Complexity of Situation	Situation's degree of complication
	Variability of Situation	The number of factors changing
Supply of Attention	Arousal of Situation	Degree of alertness/readiness for activity stimulated by the situation
	Concentration of Attention	Degree to which thoughts are brought to bear
	Division of Attention	Ability to spread or distribute focus of attention
	Spare Mental Capacity	Mental ability available for new variables
Situational Understandin g	Information Quantity	Amount of knowledge received and understood
	Information Quality	Goodness or value of knowledge communicated
	Familiarity	Degree of prior situation experience and knowledge

NASA TLX Scores

For the NASA TLX scores, participants are asked to rate the degree of burden imposed by mental demand, physical demand, time pressure, effort required, pressure due to performance level and level of frustration on a scale from 1 to 7. These ratings are then combined in a weighted average to give a TLX value. The weights assigned are determined by counting the number of times each dimension was judged by each respondent to be a greater contributor to task burden than other dimensions it was compared with. This number, divided by 15, becomes the weight used to combine the rating values to obtain a TLX value between 1 and 7. In this experiment, pair-wise comparisons were completed twice: once on the first day and once on the last day. The weights obtained were averaged between the two days, and these values were used to combine individual ratings into a TLX value.

Statistical Method Used

The fact that the SAGAT, SART and TLX source data were not based on equal interval scales, that the size of the data sets were small, and that these data sets were demonstrably non-normal in distribution, the use of *t*-tests in data analysis was ruled out. Instead, all analysis was done through pair-wise Mann-Whitney U tests. Briefly, in situations where n ranks are distributed between two data sets A & B of sample sizes n_A and n_B, it can be shown that the sampling distributions of T_A and T_B (sum of ranks) closely approximates the form of a normal distribution, provided that n_A and n_B are both equal to or greater than 5. The U statistic for each data set (A or B) is defined to be the maximum possible sum of ranks minus the observed rank sum value (T_A or T_B**)**. This is essentially the same as ordering the data from two sets, transforming the data into rank numbers and doing parametric *t*-tests on the rank number distributions.

The null hypotheses for the experiment were the following:

1. The SAGAT Level 1, 2 and 3 SA scores with radar support are no better than those given by the baseline, i.e. no radar;
2. Each of the five radar support option ATCO SAGAT Level 1, 2, and 3 SA scores are no different than those of any other radar support option;
3. The 3-D SART ratings with radar support are no better than those given by the baseline;
4. Each of the five radar support options results in ATCO 3-D SART L ratings are no different than those of any other radar support option; and
5. The NASA-TLX scores with each radar support option are no different from those given by the baseline or any other radar support option.

RESULTS

Learning Curve/Maturation

There were concerns that the data gathered during the test would show evidence of a learning curve effect in which SA scores in initial trials were worse than subsequent scores. However, a plot of SAGAT Level 1 SA score totals, shown in Figure 3 below, shows no such correlation.

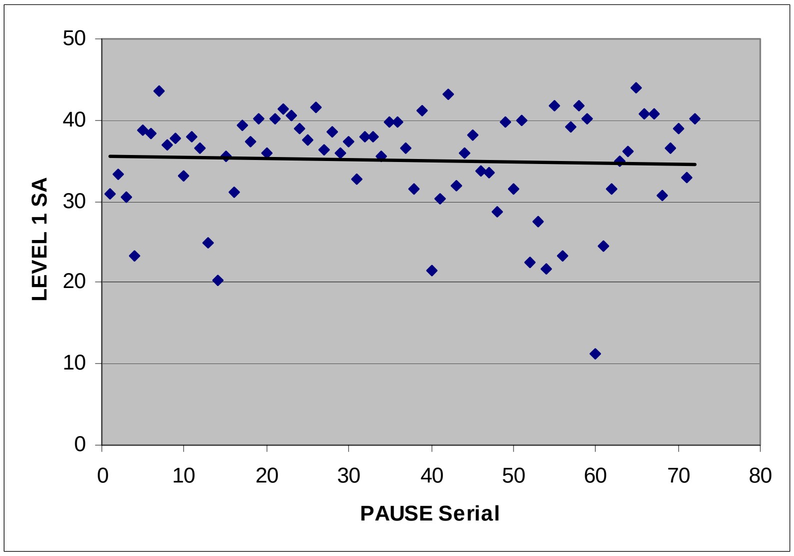

Figure 3. Sum of Level 1 SA items by pause with linear fit

SAGAT Scores

Figure 4 shows the mean and standard deviation of the ATCO Level 1 SA scores for each radar support option using three different sets of weights. The first of each bar gives the Level 1 SA scores when using a simple sum of the component SA scores. The second is computed using the equal player weights and the third using the Tower-dominant weights (i.e. ATCO-dominant weights).

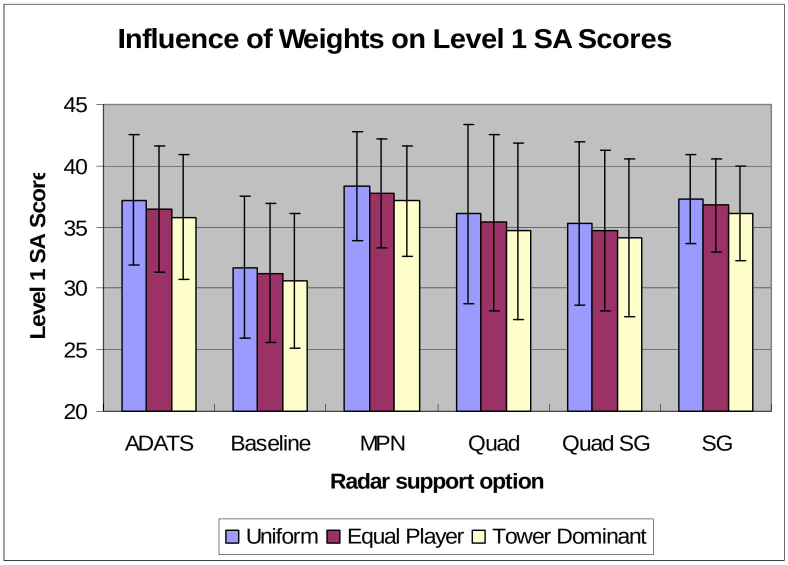

Figure 4. ATCO Level 1 SA Scores for three Weightings

One-tail Mann-Whitney U tests determined confidence levels for the improvement in ATCO SA due to radar support over baseline levels. These tests showed that all radar options resulted in significantly better Level 1 SA than the Baseline, at the 90% confidence level, with no significant differences between different radar support options.

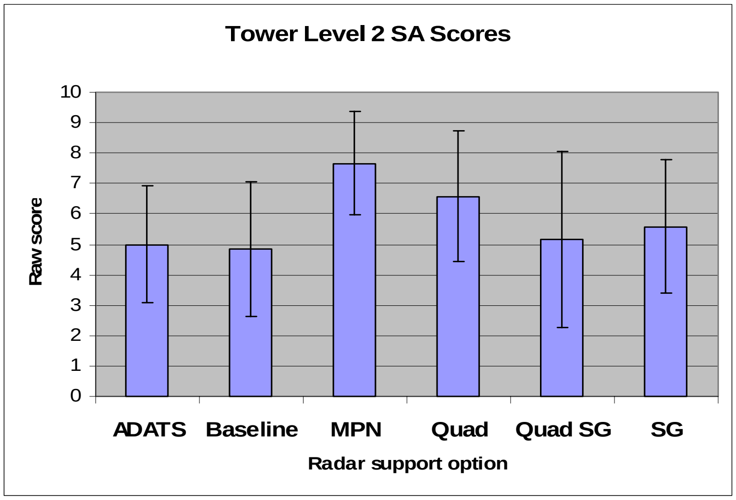

Figure 5. Raw ATCO Level 2 SA Scores

Only the MPN and Quad ATC radar options resulted in significantly better Level 2 SA than the Baseline at the 95% confidence level. Level 2 SA scores also show that MPN is significantly different and better than the other radar support options.

Table 3. *Confidence Levels that ATCO Level 3 SA Values with Radar Support Differ from Baseline*

Radar Options	2 minutes	5 minutes	Combined 2 & 5 Minutes
ADATS	69.8%	74.7%	73.7%
MPN	**93.0%**	73.7%	87.0%
QUAD	59.1%	37.5%	50.0%
QUAD/SG	52.3%	5.6%	19.3%
SG	62.5%	66.7%	70.8%

Only the MPN radar option shows statistically significant improvement in Level 3 SA over that provided by the Baseline, and only for the 2-minute future time horizon. The QUAD/Skyguard combination shows a significantly different and worse Level 3 SA than any other radar option for the 5-minute future time horizon, indicating that two different types of radars in different locations passing information in different formats may reduce Level 3 SA in the Tower. ATC radar operators use degrees and nautical miles while air defence radar operators pass information in mils and kilometres.

SART Scores

The 3-D SART showed no statistically significant differences in the demands on ATCO attention between the six variations in radar support. However, the MPN, Quad and QUAD/Skyguard options were all rated significantly better than the Baseline option for situational understanding. In addition, the MPN scores for situational understanding were significantly different (two-tail test) and better than all other radar support options, at a greater than 90% confidence level. This may reflect the ATCO's comfort with a display in the tower and the terminology used by the supporting ATC-trained personnel operating the system.

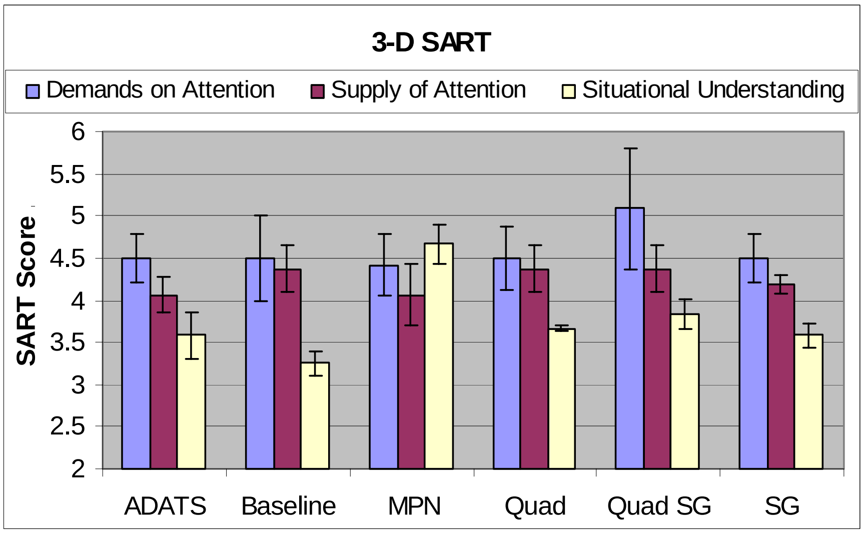

Figure 6. SART Scores for Demand on Attention, Situational Understanding and Supply of Attention

TLX Scores

An analysis of the 24 TLX data points from the ATCO indicate that the degree of burden for the ATCO was significantly different (Mann-Whitney two-tail test) and lower for the MPN radar than for any other option, at a better than 97% confidence level. All other radar options showed no significant differences between their TLX distributions from the baseline.

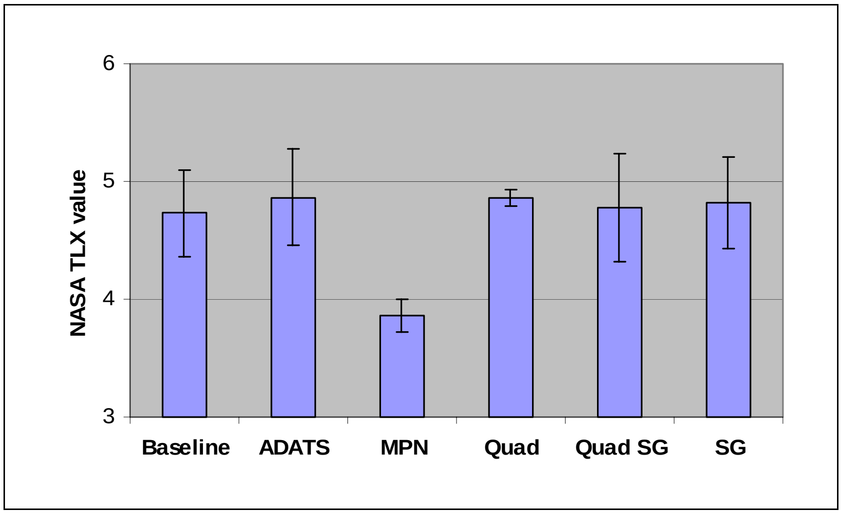

Figure 7. ATCO NASA-TLX Scores

Information display options

Based on observations of increases in ATCO SA during the MPN-25 scenarios, three information display options were examined to look at situational awareness aids in the tower. In three scenarios, the ATCO was provided with:

1. A moving map display,
2. A moving map display and MPN-25 feed, or
3. An Air Defence System Integrator (ADSI) display.

The ATCO's confidence in correctly representing the location of the TUAV was increased by having an SA aid for reference. When paired with the MPN-25 feed, the moving map display was used less; it was mostly used when the MPN-25 lost the TUAV due to terrain masking but was referred to occasionally for confirmation. When the ADSI was employed, the moving map display was removed from the tower, as its information was superfluous. Due to familiarity with standard radar displays and terminology, the ATCO stated that he was most comfortable with the MPN-25 display, however adapted quickly to the ADSI display and highly appreciated the utility of integrating the various radar pictures. The benefit of the ADSI was seen in the rapid updating of the ATCO's mental picture of the airspace (reducing errors and increasing confidence). This allowed the ATCO to give incoming aircraft relative bearings (clock face method) and distances with minimal cognitive effort compared to making calculations from his mental representation.

CONCLUSION

The employment of any radar in support of ATC operations significantly improved the Level 1 SA of the ATCO in the experiment. Only the MPN-25 radar option shows statistically significant improvements in Level 2 and Level 3 ATCO SA. This seems to indicate that ATC radars are better than AD systems in providing the ATCO with SA information. The MPN-25 SART scores for situational understanding and the NASA-TLX scores for workload were significantly better than all other radar support options. It appeared that the decrease in workload associated with the provision of an SA aid (display) allowed the ATCO to better leverage his SA into SU. Based on these findings, and considering the limitations of the simulation, the MPN-25 provided the best radar support option, but any of the radars examined in this experiment would assist in the establishment of positive control over TUAV operations in the KIA airspace.

Three aids supporting rapid report generation were the advances made in analysis methodologies, the data capture techniques employing shared workbooks, and the automation of repetitive and labour-intensive analysis tasks. The Mann-Whitney U test for statistical significance was useful even when there were only four points in each data set. However, more runs should be planned for any experiment if significance is in doubt.

The use of SAGAT, SART and NASA-TLX proved effective in determining differences in SA and task burden. The use of simulation to create an immersive synthetic environment with sufficient contextual cues to elicit expert performance was found to be a highly effective and cost efficient means of gaining insight into fielding and employment issues for new mission equipment. The combination of both rigorous measurement techniques and quality synthetic environments exceeded expectations and provided useful feedback to commanders in the field.

REFERENCES

Denford, Maj J.S., Steele J., Kalantzis E., and Roy R.L. 2003. *Limited Objective Experiment 0301 Tactical UAV Operations in Controlled Airspace*, Report AEC-R 0303, October 2003.

Endsley, M. R. 1987. SAGAT: A methodology for the measurement of situation awareness. NOR DOC 87-83, Hawthorne, CA: Northrop Corporation.

Endsley, M. R. 1988. "Design and evaluation for situation awareness enhancement". In *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 97-101). Santa Monica, CA: Human Factors Society.

Endsley, M.R. 1995. "Measurement of Situation Awareness in Dynamic Systems," *Human Factors*, 37(1), 65-84.

Endsley, M. R. 2000. "Direct measurement of situation awareness: validity and use of SAGAT." In *Situation Awareness Analysis and Measurement*. (Endsley, M. R. and Garland D. J. eds.) Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R., and Rodgers, M. D. 1994. *Situation awareness information requirements for en route air traffic control*. DOT/FAA/AM-94/27. Washington, D.C.: Federal Aviation Administration Office of Aviation Medicine.

Hart, S. G. 1986. *NASA Task Load Index (TLX) V1.0 Users Manual*. Hanover, MD: NASA Center for Aerospace Information (CASI).

Steele, J; Roy, RL; Kalantzis, E; Denford, JS. 2004. *Measuring Air Traffic Control Situational Awareness: Analysis of an Experiment Evaluating TUAV Radar Options in the Kabul Area of Operations*. Defence Research and Development Canada – Centre for Operational Research and Analysis.

Taylor, R.M. (1989) "Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design". In *Situational Awareness in Aerospace Operations*, AGARD-CP-478. 3/1 - 3/17. Seuilly-sur-Seine, France: NATO-AGARD.

Uhlarik, J. and Comerford, D. A. (2002). "A Review of Situation Awareness Literature Relevant to Pilot Surveillance Functions." (DOT/FAA/AM-02/3) Washington, DC: Federal Aviation Administration, U.S. Department of Transportation.

Annex A – Level 1 SA

Players were asked to compare the importance of each level 1 item to airspace safety. A value of 1 for each Yes and a value of -1 for each No were entered to determine the priority function of each Subject Matter Expert (SME). The table below gives uniformly averaged values and non-uniformly averaged values used to compute the Equal Player Weights and the ATCO Dominant Weights, respectively. The non-uniform average gave 50% influence to the ATCO's responses and 12.5% to each of the four other players.

Table 4. *The calculation of Equal Player Weights and ATCO Dominant Weights*

	AC TYPE	ALTITUDE	CALL SIGN	HEADING	LOCATION	SPEED	ALT CHANGE	HDG CHANGE	ACTIVITY	EMERGENCY
Equal Player Row sums:	-5.4	1.0	-7.4	1.0	5.8	-1.4	3.4	0.2	-2.2	5.0
Uniform Weights:	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Equal Player Influence:	-0.54	+0.10	-0.74	+0.10	+0.58	-0.14	+0.34	+0.02	-0.22	+0.50
Equal player Weights:	0.46	1.10	0.26	1.10	1.58	0.86	1.34	1.02	0.78	1.50
ATCO Dom. Row sums:	-4.0	+2.5	-7.25	+2.5	+7.0	-2.75	+3.25	-0.25	-2.5	+3.25
Uniform Weights:	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
ATCO Dom. Influence:	-0.40	+0.25	-0.725	+0.25	+0.70	-0.275	+0.325	-0.025	-0.25	+0.35
ATCO Dom. Weights:	0.40	1.25	0.275	1.25	1.70	0.725	1.325	0.975	0.75	1.35

Table 5. *Trial results for Level 1 SA Scores assuming all items are equally important to air safety*

PAUSE	ADATS	BASELINE	MPN	QUAD	QUAD SG	SG
1. A	43.6	31.8	29.8	23.29	24.8	33.67
1. B	35.57	39.25	43.67	36.4	37.5	33.4
1. C	33.8	23.14	32.5	21.57	36.5	39.75
2. A	44	23.5	37	40.5	39.5	42.17
2. B	36.5	32.86	32	42.75	20	32.8
2. C	40	37	41.6	38.75	42.25	37.75
3. A	26.67	28.29	38.8	30.14	31.43	39.33
3. B	39.4	31.43	43.5	34	32.8	31.33
3. C	37.25	35.75	40	39	36.8	36
4. A	43	38.25	39	42.25	39	38.25
4. B	29.75	23	40.4	45.75	42.2	42.4
4. C		36.5	41.6	38.4	40.71	40.25
Mean	37.23	31.73	38.32	36.07	35.29	37.26
Std Dev	5.35	5.75	4.41	7.25	6.65	3.60

Table 6. *Trial results for Level 1 SA Scores: Equal Player Weights (assumes that all players are equally expert in air safety issues)*

PAUSE	ADATS	BASELINE	MPN	QUAD	QUAD SG	SG
1. A	43.18	30.972	29.588	22.82	24.264	33.2733
1. B	34.89	37.995	42.847	33.992	36.75	32.18
1. C	32.83	22.543	31.743	21.3457	35.55	39.225
2. A	43.71	23.97	36.767	39.75	39.0167	42.07
2. B	34.69	31.274	30.85	41.945	19.71	32.48
2. C	38.82	37.116	40.616	38.395	41.515	37.605
3. A	27.48	27.811	38.744	29.6543	31.28	38.24
3. B	38.58	31.246	43.27	33.576	32.48	30.5133
3. C	35.11	34.305	39.88	37.5933	36.336	35.69
4. A	42.41	39.065	37.9	42.005	38.19	36.905
4. B	29.68	22.946	38.848	45.325	41.484	42.528
4. C		35.82	41.68	37.968	40.2029	40.615
Mean	36.49	31.26	37.73	35.36	34.73	36.78
Std Dev	5.14	5.63	4.45	7.18	6.53	3.84

Table 7. *Trial results for Level 1 SA Scores: ATCO Dominant Weights*

PAUSE	ADATS	BASELINE	MPN	QUAD	QUAD SG	SG
1. A	42.66	30.24	29.08	22.379	23.865	32.73
1. B	34.16	36.994	42.24	32.45	36.2625	31.7
1. C	32.12	22.104	31.02	20.621	34.95	38.53
2. A	43.24	23.475	36.37	38.762	38.104	41.425
2. B	33.64	30.3786	30.14	41.112	19.144	32.015
2. C	37.87	36.39	39.91	37.731	40.825	37.219
3. A	27.3	27.229	38.06	29.264	30.886	37.075
3. B	37.94	30.886	42.87	33.135	32.015	29.683
3. C	33.82	34.05	39.4	36.958	35.535	35.319
4. A	41.62	38.356	37.3	41.506	37.4438	35.75
4. B	29.28	22.457	37.92	44.987	41.01	42.03
4. C		34.987	41.2	37.11	39.5107	40.319
Mean	35.79	30.629	37.124	34.668	34.129	36.150
Std Dev	5.09	5.527	4.488	7.170	6.455	3.859

Annex B – ATCO Level 2 SA

Level 2 SA is the awareness of the practical consequences of Level 1 SA. In this experiment, it was measured by asking the ATCO to indicate in order of proximity the three aircraft that were currently closest to the UAV and scoring the responses. Scores include 6 points for the closest aircraft, 3 for the next closest and 1 for the third closest for a score out of 10.

Table 8. *ATCO Level 2 SA results.*

PAUSE	ADATS	BASELINE	MPN	QUAD	QUAD SG	SG
1. A	6	2	9	7	1	6
1. B	7	6	6	3	7	8
1. C	2	8	8	4	5	4
2. A	8	3	9	4	3	2
2. B	3	3	7	8	1	5
2. C	3	2	10	6	4	4
3. A	6	3	8	10	9	9
3. B	5	6	4	7	5	3
3. C	4	5	7	10	4	8
4. A	4	8	6	8	4	4
4. B	4	8	8	6	10	8
4. C	8	4	10	6	9	6
Mean	5.00	4.83	7.67	6.58	5.17	5.58
Std Dev	1.91	2.23	1.70	2.14	2.88	2.18

Annex C – ATCO Level 3 SA

Level 3 SA is the awareness of the near-future significance of current level 1 SA. In this experiment, it was measured by asking the ATCO to indicate in order of proximity the three aircraft that will be closest to the UAV in 2 minutes time and in 5 minutes time. Scores were assigned in the same way as for level 2 SA. Results are summarized in the following tables.

Table 9. *ATCO Level 3 SA results: 2 minutes later.*

PAUSE	ADATS	BASELINE	MPN	QUAD	QUAD SG	SG
1. A	10	4	6	4	0	2
1. B	6	6	10	3	6	8
1. C	4	8	10	4	7	4
2. A	4	3	5	4	1	1
2. B	5	3	9	4	2	8
2. C	6	6	9	4	10	9
3. A	6	3	8	1	6	5
3. B	3	2	5	4	8	4
3. C	4	10	3	10	5	6
4. A	10	9	7	10	8	5
4. B	4	6	5	8	6	8
4. C	6	4	6	10	4	7
Mean	5.67	5.33	6.92	5.50	5.25	5.58
Sigma	2.17	2.49	2.18	2.99	2.89	2.43

Table 10. *ATCO Level 3 SA results: 5 minutes later.*

PAUSE	ADATS	BASELINE	MPN	QUAD	QUAD SG	SG
1. A	10	4	4	5	1	10
1. B	4	5	10	2	3	5
1. C	3	10	10	6	4	4
2. A	7	3	3	3	2	1
2. B	4	6	5	7	3	4
2. C	6	0	7	4	1	6
3. A	6	1	7	2	1	5
3. B	7	3	4	6	4	6
3. C	10	8	2	4	8	10
4. A	6	6	10	9	8	5
4. B	0	9	4	3	3	10
4. C	10	8	8	8	1	4
Mean	6.08	5.25	6.17	4.92	3.25	5.83
Sigma	2.93	3.03	2.76	2.22	2.39	2.70

Annex D – SART SCORES

This annex provides the 3-D SART results obtained from the experiments.

Table 11. *3-D and Global SART results.*

	ITERATION	ADATS	BASELINE	MPN	QUAD	QUAD SG	SG
DEMANDS:	1	4.00	4.00	4.33	4.00	4.00	4.00
	2	4.67	5.00	4.00	4.33	5.33	4.67
	3	4.67	5.00	4.33	4.67	5.00	4.67
	4	4.67	4.00	5.00	5.00	6.00	4.67
	Mean	4.50	4.50	4.42	4.50	5.08	4.50
	Std Dev	0.33	0.58	0.42	0.43	0.83	0.33
SUPPLY:	1	4.25	4.00	4.25	4.00	4.00	4.25
	2	4.25	4.5	3.5	4.75	4.75	4.25
	3	4.00	4.25	4.00	4.25	4.25	4.00
	4	3.75	4.75	4.50	4.50	4.50	4.25
	Mean	4.06	4.38	4.06	4.38	4.38	4.19
	Std Dev	0.24	0.32	0.43	0.32	0.32	0.13
UNDERSTANDING:	1	4.00	3.00	4.33	3.67	4.00	3.67
	2	3.33	3.33	4.67	3.67	3.67	3.33
	3	3.33	3.33	4.67	3.67	3.67	3.67
	4	3.67	3.33	5.00	3.67	4.00	3.67
	Mean	3.58	3.25	4.67	3.67	3.83	3.58
	Std Dev	0.32	0.17	0.27	0.00	0.19	0.17
GLOBAL SART: (= UNDERSTANDING + SUPPLY – DEMANDS)	1	4.25	3.00	4.25	3.67	4.00	3.92
	2	2.92	2.83	4.17	4.08	3.08	2.92
	3	2.67	2.58	4.33	3.25	2.92	3.00
	4	2.75	4.08	4.50	3.17	2.50	3.25
	Mean	3.15	3.13	4.31	3.54	3.13	3.27
	Std Dev	0.74	0.66	0.14	0.42	0.63	0.45

Annex E – NASA-TLX WEIGHTS and SCORES

Twice during the exercise (on Tuesday and Friday), the participants indicated the relative importance of each of these dimensions in contributing to the sense of burden. A 1 was counted for each item that was circled as being more important. The sums of these of each of the six assessment factors were entered in the following table.

Table 12. *TLX burden contribution data for Tuesday, Friday, and averaged.*

	RESPONDENT	MENTAL DEMAND	PHYSICAL DEMAND	TEMPORAL DEMAND	EFFORT	PERFORM	FRUSTRATION	TOTAL
TUESDAY	Msn Comd	5	0	3	2	4	1	15
	ATCO	5	0	2	3	1	4	15
	AD Radar 1	4	0	2	4	4	1	15
	AD Radar 2	3	0	4	2	5	1	15
	ATC Radar	1	0	3	3	3	5	15
FRIDAY	Msn Comd	4	1	3	2	5	0	15
	ATCO	5	0	4	2	2	2	15
	AD Radar 1	3	1	0	3	5	3	15
	AD Radar 2	5	0	3	2	4	1	15
	ATC Radar	5	0	4	3	2	1	15
AVERAGE	Msn Comd	4.5	0.5	3	2	4.5	0.5	15
	ATCO	5	0	3	2.5	1.5	3	15
	AD Radar 1	3.5	0.5	1	3.5	4.5	2	15
	AD Radar 2	4	0	3.5	2	4.5	1	15
	ATC Radar	3	0	3.5	3	2.5	3	15

The NASA-TLX Score is the sum of the products of the average weights above and the TLX values entered on a scale of 1 to 7 for each factor by the participants.

The following table summarizes the TLX scores from the ATCO. Physical Demand is not included because the weights given to it by the ATCO was 0.

Table 13. *ATCO TLX rating responses.*

ATCO	RUN #	BASELINE	ADATS	MPN	QUAD	QUAD SG	SG
Mental	1	4	4	4	5	4	4
Demand	2	5	5	4	5	5	5
(Weight = 5/15)	3	5	5	4	5	5	5
	4	5	5	4	5	5	5
Temporal	1	4	4	4	5	3	4
Demand	2	5	5	4	4	4	4
(Weight = 3/15)	3	5	5	4	5	5	4
	4	4	5	4	5	5	5
Effort	1	4	4	4	4	4	4
	2	5	5	3	5	5	5
(Weight = 2.5/15)	3	5	5	4	5	5	5
	4	4	6	4	5	5	6
Performance	1	4	4	5	4	4	4
	2	4	4	5	4	4	4
(Weight = 1.5/15)	3	4	4	4	4	4	4
	4	4	2	4	4	4	4
Frustration	1	6	5	3	5	5	5
	2	6	6	4	6	6	6
(Weight = 3/15)	3	6	5	4	5	6	6
	4	4	5	4	5	6	6
ATCO TLX scores	1	4.40	4.20	3.90	4.73	4.00	4.20
	2	5.10	5.10	3.93	4.90	4.90	4.90
	3	5.10	4.90	4.00	4.90	5.10	4.90
	4	4.33	4.87	4.00	4.90	5.10	5.27