ABSTRACT

Objective Situation Awareness Measurement Based on Performance Self-Evaluation

Joe De Maio

The research was conducted in support of the NASA Safe All-Weather Flight Operations for Rotorcraft (SAFOR) program. The purpose of the work was to investigate the utility of two measurement tools developed by the British Defense Evaluation Research Agency. These tools were a subjective workload assessment scale, the DRA Workload Scale and a situation awareness measurement tool. The situation awareness tool uses a comparison of the crew's self-evaluation of performance against actual performance in order to determine what information the crew attended to during the performance. These two measurement tools were evaluated in the context of a test of innovative approach to alerting the crew by way of a helmet mounted display. The situation assessment data are reported here. The performance self-evaluation metric of situation awareness was found to be highly effective. It was used to evaluate situation awareness on a tank reconnaissance task, a tactical navigation task, and a stylized task used to evaluated handling qualities. Using the self-evaluation metric, it was possible to evaluate situation awareness, without exact knowledge the relevant information in some cases and to identify information to which the crew attended or failed to attend in others.

OBJECTIVE SITUATION AWARENESS MEASUREMENT BASED ON PERFORMANCE SELF-EVALUATION

Joe De Maio
Army/NASA Rotorcraft Division
U. S. Army Aviation and Missile Command
Ames Research Center
Moffett Field, CA

The present research was performed as part of the U. S. Army Human Systems Integration and the NASA Safe All-weather Flight Operations for Rotorcraft (SAFOR) programs. The goal of this program is to make dramatic reductions in the rate and severity of civil and military rotorcraft accidents attributable to human error. A major factor in pilot error is the failure to perceive or to interpret flight related information. The term, situation awareness, includes this perception and interpretation of relevant information. Situation awareness and workload are independent but interacting factors. The present research evaluated techniques for measuring both factors, and the present report describes the situation awareness portion of the work.

The present research was also shaped by a cooperative effort between the Army Aeroflightdynamics Directorate, located at NASA Ames Research Center and the British Defense Evaluation and Research Agency (DERA). The DERA is developing a battery of workload and situation awareness measures, applicable to both U. S. and British research. The present research evaluated variants of two measurement techniques, the DRA Workload Scale (DRAWS) and self-assessment probes. DRAWS is a four-dimensional, subjective rating of perceived workload. The self-assessment probes are performance self-evaluations in which the individual rates his own performance against pre-determined criteria. The accuracy of the performance self-evaluation provides a measure of situation awareness.

There has been much discussion regarding the exact meaning of the term situation awareness (McMillan, Bushman, and Judge, 1996). Several approaches have been proposed to measure the construct. Fracker (1991) distinguished between two broad classes of situation awareness measures. Explicit measures require the operator to recall facts relevant to the performance of the task, that is, the operator tells the evaluator explicitly what he knows about the task. In implicit measurement, the evaluator measures task performance and infers a level of operator situation awareness from performance. The logic here is that if the operator were aware of a certain fact, he would perform a certain action. If he fails to perform the action, then he must have been unaware of the relevant fact.

There are difficulties associated with both implicit and explicit measurement of situation awareness. The logic of implicit measurement is sound, that is, if A implies B, then not B implies not A. It is not easy to be certain that knowledge of a given fact will lead inevitably to a specific behavior, that is, that the premise is in fact true. The more complex the knowledge and the behavior, the more difficult it becomes to be certain of the linkage between them. As a result, implicit measures have been used for only simple facts and responses (e.g., Eubanks and Killeen, 1983). Even when the logical requirements of implicit measures are met, there can still be problems with underlying statistical assumptions. For example, Eubanks and Killeen assumed normality and equal variance of signal and non-signal in order to apply the theory of Signal Detectability to a targeting task. Long and Waag (1981) have pointed out that this assumption is wrong for many supra-threshold tasks where situation awareness might be a concern.

While explicit measures do not require this sine qua non relationship with performance, the issue of relevance of facts remains. In explicit measurement, the evaluator queries the operator about specific facts determined a priori to be relevant to the task. The risk here is that the importance of a particular fact may vary over the course of task execution. Evaluations of situation awareness may differ depending on the part of the task queried using that fact (Fracker, op cit). For example, Endsley (1995b) used altitude as a query subject in an evaluation of the situation awareness global assessment technique (SAGAT). The flight task was a combat air patrol. This mission can be broken into two major components, orbit and air combat. During orbit the pilot is to maintain a constant altitude and monitors the altimeter continually. During air combat, maintaining a specific altitude is relatively unimportant, and the pilot focuses more on control variables to achieve a tactical

advantage. As a result Endsley found very high variability, with evidence of good situation awareness (accurate recall of altitude) mixed with virtually non-existent situation awareness (errors over 20,000 ft). This risk may be minimized by using a battery of well chosen queries, from which inappropriate variables can be culled.

Explicit measures of situation awareness have proven most popular because of their versatility. Explicit measures can address not only raw facts (e.g., altitude), but also state concepts consisting of an aggregation of facts (e.g., tactical advantage) and also future states (e.g., engagement outcome). Endsley (1995a) has labeled raw facts, aggregated state concepts and future states situation awareness levels one, two, and three respectively.

Because explicit measures require the operator to respond to a query, there is always a memory component to the response. The memory component can be minimized by halting the task and querying the operator about his state immediately prior to the halt. This has been called a concurrent memory probe (Fracker, op cit). The liability of the concurrent probe is that the task must be suspended or terminated to allow the query. This is not always feasible, and even when it is, the halt can severely disrupt the performance of the task.

Fracker (op cit) has suggested that level two and three factors may persist long enough to be probed retroactively, that is following normal task completion. Endsley (1995b) has tested the effect of delaying report by querying a number of facts following task halt. She has shown that even level one concurrent memory probes can be stable when many facts must be recalled.

The present research evaluated a form of retroactive, level two query. The self-evaluation required the pilots to integrate a number of facts about their performance in order to produce an evaluation. Our queries differed from those generally applied. Whereas level two queries generally require the aggregation of a defined set of facts available to the operator, our self-evaluations queried the overall performance of a task without asking for values of specific variables at specific times. In some instances the ultimate performance data for grading the task was not even available to the pilot. We determined whether the pilot was aware of an aspect of the task by comparing the self-evaluations with objective evaluations of performance.

The evaluation of situation awareness metrics was done in a test of helmet mounted display symbology. We tested two types of symbology, navigation aids and alerts. We present the results of this test elsewhere (De Maio and Hart, in preparation). The mission tasking was designed around the symbology evaluation, and the situation awareness data were gathered at appropriate times. Our concept of situation awareness was that it entails awareness of those aspects of the task and the environment that affect the quality of performance. Therefore a situationally aware pilots will show an ability to evaluate their own performance regardless of how good that performance is. We tested this concept by comparing pilots' self-evaluation of performance with actual performance in a variety of tasks.

METHOD

Apparatus

Helicopter Simulation - The simulator was the six-degree-of-freedom Vertical Motion Simulator (VMS). The VMS is unique among flight simulators in its large range of motion to provide flight cues to the pilot. A rotorcraft cabin was configured as a single-pilot cockpit with a four-window computer generated display, having three forward view, CRTs, (27° X 147°) and one CRT chin window on the right side (26° X 22°). Out-the-window imagery was generated by an Evans and Sutherland ESIG 3000 image generator. The simulated aircraft was a UH-60A. Rotor, engine, and transmission sounds were simulated. Conventional helicopter controls were used. The visual throughput time delay was approximately 72 msec.

Panel instruments were displayed on two 14 in. color CRTs. The right CRT displayed generic flight instruments, and the left CRT displayed a moving map of the visual data base (see Figure 1). The map showed major terrain features and roads in light blue. The planned course was in red. A compass rose was in the upper right hand corner. A gray square overlaid the high resolution area at the center of the data base. In the visual data base, this was a high definition rendering of a small village, that could not be represented on the map. A digital range indicator in the lower left corner indicated the size of the displayed area in nautical miles. When the alert task was presented, the required input and the pilot's response were overlaid on the bottom of the map. The helmet mounted display system

consisted of the helmet, helmet display unit and head position sensing system of the AH-64 integrated helmet and display sight system.

Helmet Display Symbology - Helmet display symbology was based on the AH-64 pilot night vision system (PNVS), cruise mode symbology. This included a compressed 120° compass at the top of the display, digital torque and airspeed on the left, digital altitude and analog, "radar" altitude and vertical speed on the right (see Figure 2a). Altitude was above ground level for both digital and analog displays. A dashed line gave a rough indication of pitch and roll, referenced to the display frame. A diamond indicating the position of the aircraft's nose is the only head slaved PNVS symbology.

Alert Task Symbology – There were five alert type conditions. A No-alert condition provided a flying performance baseline. Alert flash (Localized and Full-Screen) was crossed with alert information content (No-Info and Partial-Info) to yield four experimental conditions. The alert was presented on the helmet mounted display, with a digit entry task to simulate the procedural response presented on the panel.

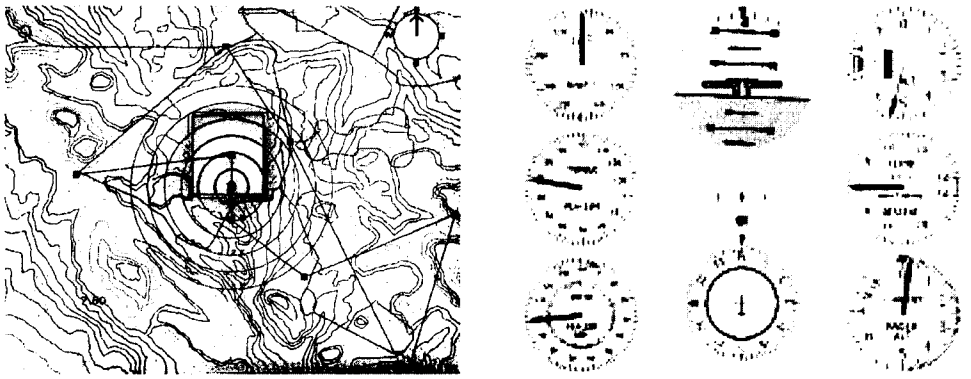
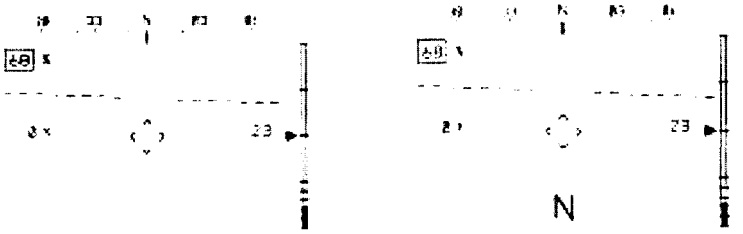


Figure 1. Simulated Instrument Panel and Map Display. Instruments Consisted of White and Colored Graphics on a Black Background. Terrain Contours Consisted of Colored Lines on a Black Background. Roads Were Teal. Planned Route Was Red. Color Scheme has been Revised for Better Printing.

In all Alert Task conditions, a flashing letter was presented in the bottom center of the display. The TADS field of view box was omitted (see Figure 2b). In the Localized alert condition, this symbol constituted the entire alert. In the Full-Screen alert condition, all symbology flashed. Three alert symbols were used. In the No-Information condition, an upper case "N" indicated an alert. In the Part-Information condition, an "L" indicated an alert and that numbers for the "procedural" task were to be entered left-to-right, while and "R" indicated right-to-left entry.

Navigation Symbology – There were three navigation display conditions. A "Visual" navigation condition used the basic PNVS symbology without the waypoint symbol. A "Waypoint" condition used a waypoint marker overlaid on the visual scene. The "CDI" condition incorporated symbols into the compass display indicating bearing to waypoints and course deviation. These two displays also included an arrival time clock in the upper right that showed the pilot's instantaneous arrival time error, up to +/-99 sec. Arrival time error was simply the difference between the target arrival time and the arrival time computed from current speed and distance remaining. In an actual mission speed would vary on each navigation leg, and so arrival time would be needed for each leg. In the simulation, planned speed was constant across legs, so only segment arrival time was displayed.

Waypoint Symbology - The Waypoint symbology consisted simply of a pennant displayed at the geographical location of each waypoint and the altitude of the aircraft. The pennants were maintained as moving models by the image generation system but were displayed by the Silicon Graphics computer that drove the helmet display. Each pennant was shaped like an arrow that pointed toward the next waypoint (see Figure 2c). Because the pennants were maintained as part of the visual data base, all were displayed continuously, and their size decreased with range.



a. Basic Symbology b. Alert Symbol

c. "Waypoint" Symbol d. "CDI" Symbology.

Figure 2. Helmet Display Symbology.

CDI Symbology - The CDI symbology mimicked a conventional, panel mounted, course director indicator (CDI) (see Figure 2d). A tail beneath the compass lubber line pivoted to point toward the planned course (no course error shown). A carat (^) showed the heading to the current waypoint (as in the PNVS). A circle (o) indicated the heading to the next waypoint. Both waypoint symbols edge limited. In figure 2 the current waypoint symbol edge limited to show only one leg of the carat (^).

Alert Task Operator Interface – The pilot reacted to the alert symbol on the helmet mounted display by performing a task overlaid on the map display using a keypad. The keypad was located on a console next to the collective lever and contained a button to allow the pilot to acknowledge the alert along with a 10-key numeric pad. The pilot acknowledged the alert and entered five random digits, left-to-right or right-to-left, as directed As the pilot entered each correct digit, it was displayed in the correct blank. Incorrect entries were ignored. Once the pilot entered the fifth correct digit, both the map and helmet mounted displays returned to the nominal state.

Experimental Tasks – A standardized flying task was developed, that consisted of segments. These were (1) take-off and cross-country navigation with reconnaissance for tanks, (2) low altitude track following, (3) cross-country flight with reconnaissance, (4) bob-up, and (5) return to and landing in the village. The simulator detected when the aircraft passed within 1000 ft of each waypoint and automatically advanced to the next segment. Data were collected on segments one through four (See Table 1.

Table 1 Pilot Tasking and Situation Awareness Ratings.

Pilot Task/ Segment	Situation Awareness Rating
Take off / 1	No
Navigation / 2	Yes – Navigation
Navigation & Reconnaissance / 3	Yes – Navigation & Reconnaissance
Track Following / 4	No
Navigation & Reconnaissance / 3	Yes – Navigation & Reconnaissance
Bob-up / 5	Yes
Approach & Land / 6	No

Twelve missions were created by varying the waypoints and tank laydowns on the two navigation segments. The take-off, track following and return segments were the same on each mission. The

number of waypoints on each segment was also constant, but their location varied. The pilot took off and flew to the first waypoint, a beacon just outside the village. From there he turned to the first of two variable waypoints. From the second variable waypoint, the pilot flew to the track. After the track following task, the pilot turned to the first of two variable waypoints on the second navigation segment, ending at the bob-up site. Following the bob-up, the pilot returned to the village by a constant route and landed.

Navigation and Reconnaissance Tasks – On each of the two cross-country segments the pilot flew from a constant location waypoint (the beacon or end of the track) to an ending point (the start of the track or the bob-up) by way of two intermediate waypoints whose location varied from mission to mission. He was to reconnoiter for tanks, whose location and number varied. The reconnaissance task was intended to increase workload by providing additional tasking and to make the navigation task more challenging by forcing the pilot off the course. When he completed his reconnaissance, he depressed the microphone switch on the cyclic grip to mark the report time and reported the number of tanks seen, over an open microphone. He was given no feedback regarding his performance on any part of this task, save the arrival time clock on the test navigation displays. In preliminary runs with unlimited visibility and a moving map display, we found navigation to be very easy. Therefore, we reduced visibility to 5000 ft and rendered the map stationary. A single alert was presented during each navigation segment. The duration of each navigation segment was 10 to 15 minutes.

Track Following Task – In the track following task, the pilot was to follow the centerline of a two-lane road at nap-of-the-earth altitude over rolling terrain. At an assigned airspeed of 40 kt, this task took about six minutes. An alert was programmed to occur randomly during each consecutive one-minute interval. If the pilot flew too fast, later alerts would not occur.

Bob-up – The bob-up task was developed from the Aeronautical Design Standard-33 (ADS-33, 1994) bob-up task used in handling qualities evaluations. In our task the pilot was to hover 10 ft above the ground, to ascend rapidly to 50 ft above ground level, to hover for 10 sec, and to descend to the low hover. Out-the-window cues to altitude and position were provided by hover boards in front of the aircraft and walls off the right nose (see 3). The alert task occurred three times in the bob-up, making the bob-up very different from the ADS-33 task. The visual and motor workload were very much higher due to the requirement for precision flying combined with the complicated alert response. The alerts were keyed to phases of the bob-up maneuver. One alert was programmed during the ascent, one during the high hover, and one during the descent. These phases were very short in duration, so the pace of activity was very rapid, and timing was critical. Total task duration was less than 30 seconds.

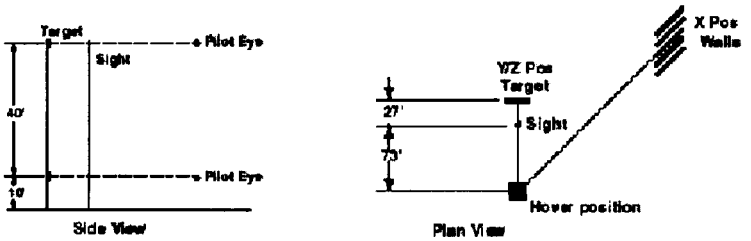


Figure 3. Hover Boards and Walls to Provide Position Cues for Bob-up Task.

Procedure

Two pilots participated in each of three week-long test periods. Each pilot performed one to three missions and then took a break while the other pilot flew. The duration of each pair's simulation period was four days. Missions lasted about 30 minutes. Following each mission the pilot gave his workload and performance self-evaluation ratings without receiving any feedback on his performance. Pilots received written instructions explaining the goal of the research and the tasks that they would perform. They then performed familiarization flights until they were ready to begin practicing the experimental tasks.

Pilots received paper maps that duplicated the cockpit map in order to familiarize themselves with the mission before hand. They were also allowed to make a list of the waypoints for each mission to take into the cockpit. As the pilot passed each waypoint, he was to depress the microphone switch and state the waypoint name. Three practice missions were provided. The pilots did not perform

reconnaissance. The pilots practiced in each navigation display and alert condition until they and the experimenter felt that they were ready for the experimental trials. Pilots gave no ratings of the practice runs.

Navigation display, alert type and alert information were presented in a full factorial design. There were three levels of navigation display, visual, CDI, and waypoint. There were two levels of alert type, localized and full-screen, and two levels of alert information, none, and partial. This gave 12 experimental conditions.

The order of presentation of the navigation display conditions for data collection was balanced by a Latin Square. Presentation of alert conditions was balanced for first order effects. Pilots flew three to six experimental missions per day for a total of 12 experimental missions.

Data collected included performance self-evaluations, mission time of waypoint passage reports, mission time of reconnaissance reports, reconnaissance reports, and automatically recorded flight performance data. The pilots gave workload ratings and performance self-evaluations orally over the intercom, and the experimenter transcribed them into a log, at the end of each mission. Data were also collected on responding to the alerts. These data are reported in a separate report. Alert response performance was not the subject of workload and situation awareness ratings.

The performance self-evaluations were based on defined error criteria for each task rated. Task performance ratings used three ordinal level values similar to those used in handling qualities evaluation. "Desired" meant the highest level of performance. "Adequate" meant performance that was acceptable but not of the highest level. "Outside of Acceptable" meant unacceptable performance. Error scores defining each performance level are presented in Table 2.

Table 2. Error Criteria for Performance Self-Evaluation Ratings.

		D (Desired)	A (Acceptable)	O (Outside of Acceptable)
Recon	Accuracy (% tanks detected)	>90%	75% - 90%	<75%
	Timeliness (report time after first detection	<20 sec	20 - 40 sec	>40 sec
Naviga-tion	Accuracy (max course deviation)	<100 ft	100 - 200 ft	>200 ft
	Timeliness (at track and bob-up)	<+/- 10 sec of assigned time	<+/- 20 sec of assigned time	>+/- 20 sec of assigned time
Bob-up	Height	<+/- 3 ft	<+/- 6 ft	>+/- 6 ft
	Time	<+/- 4 sec	<+/- 6 sec	>+/- 6 sec
	Position	<+/- 6 ft	<+/- 10 ft	>+/- 10 ft

RESULTS

Reconnaissance Situation Awareness

The reconnaissance task performed several functions in the simulation. It served as a workload enhancer in its own right, and it increased the difficulty of the navigation task by forcing the pilot to deviate from the planned course in search of the targets. It also provided a way of testing the performance self-evaluation situation awareness metric. In some sense reconnaissance performance is the most interesting application of the technique because the pilots have no feedback about the correctness of their report. So pilots must base their evaluation on their perception of the situation when they performed the task.

Following each simulated flight, the pilot provided a rating of the quality of his performance on the reconnaissance task. A single rating covered two performances, one on the first navigation segment and one on the second. We compared this rating with one made by the experimenter using the

actual detection data. Scoring of the detection data was complicated by the fact that the pilots sometimes wanted to report individual tanks and sometimes wanted to report "platoons." We accommodated the pilots' desire to report platoons by devising a scheme for converting the number of tanks observed to number of platoons. A platoon consists of four tanks, so the pilots were told to consider four or fewer tanks to be one platoon. When a pilot found more than four tanks, he was to consider groups of four to be platoons and any remaining tanks to be elements of a platoon. In effect then the pilot divided the number of tanks by four and rounded any remainder up.

Table 3 shows a comparison of pilots' detection performance with their self-evaluation accuracy. Objective criteria were needed in order to evaluate reconnaissance performance. Criteria were provided to the pilots for "Desired," (D) "Adequate," (A) and "Outside of adequate" (O) performance. The criteria against which the pilots scored themselves had only addressed missed tanks, but pilots also reported too many tanks. So the experimenter's criteria used percentage deviation from actual without regard to the sign of the deviation. So reporting one tank too many was the same as reporting one tank too few.

Table 3. Comparison of Pilots' Reconnaissance Performance Self-Evaluation with Actual Performance.

	Pilot 1	Pilot 2	Pilot 3	Pilot 4	Pilot 5	Pilot 6
Proportion Correct Detections	0.38	0.54	0.86	0.66	0.64	0.71
Proportion Correct Self Evaluations	0.08	0.30	0.67	0.17	0.45	0.25

Reports of platoon and individual tanks were combined by multiplying the deviation from the correct number of platoons by four and treating the result as a deviation in the number of tanks reported. So one platoon was converted to four tanks. The pilots showed a substantial range of performance in detection of the tanks. The worst pilot detected fewer than 40% of the targets, while the best detected up to more than 80%. The pilots' self-evaluations showed an even greater range, and they were highly correlated with performance ($r = 0.78$, $p < 0.07$ (Beyer, 1966); see Figure 4). On the whole the pilots missed a substantial number of the tanks. They also showed a modest level of situation awareness, with an average self-evaluation accuracy of 0.32. This result may actually support the conclusion that pilots who were less effective in the reconnaissance task were in fact less situationally aware. That is, they were less able to tell when they had performed a thorough search, while the more effective pilots were able to tell how thoroughly they had searched, even without the benefit of feedback on their effectiveness. Pilots 3 and 5 showed fair accuracy in their self evaluations, even though they had a substantial rate of miss reporting of tanks. The performance self-evaluation does appear to be a good measure of situation awareness.

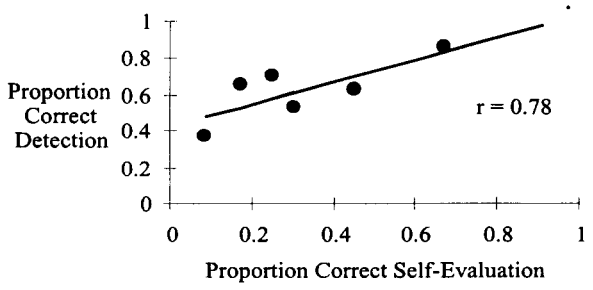


Figure 4. Relationship between Self -Evaluation and Actual Performance. Correlation is Significant at $p < 0.07$ (Beyer, op cit).

Bob-Up Situation Awareness

We encountered many technical problems with the bob-up task. These stemmed from the difficulty of responding to the alerts while maintaining aircraft control. The alert task required the pilots to remove their hand from the collective, which had a severe impact on aircraft control. The timerequired to perform the alert response task was so long that the response to an alert in one phase of the bob-up (e.g., ascent) was not completed until a subsequent phase (e.g., high hover). This caused problems with the timing of data collection. We were able to get self-evaluation data which shed some light on how the pilots reacted to an overwhelming task ensemble. We shall examine three aspects of bob-up performance: time in the high hover, maximum altitude error in the high hover, and maximum horizontal position error.

The pilots found this task ensemble nearly impossible (see Table 4). They met the time criteria for holding the high hover on fewer than 50% of the trials. They met altitude criteria on only about 15% of the trials, and none succeeded in meeting the horizontal position criteria. Performance this poor raises concern about a flaw in the simulation, but we checked the flight model and hardware thoroughly and found none. When we examined the self-evaluations, the reason for the poor performance became apparent. The pilots simply neglected the precision hover task (presumably to perform the alert response). This task had been included to examine the validity of very frequent alerts (which would provide a larger amount of data). In the event, the task was too unrealistic. Normally the pilot would simply have landed and performed the alert task on the ground.

The pilots were moderately accurate in the self-evaluations of time in hover. Their self-evaluations were correct on 62% of the trials. Their accuracy was low for performance in the "Desired" category. They showed the best accuracy when they rated their performance "Outside of Acceptable." Errors of up to several hundred percent of the desired let them achieve 92% accuracy. This compares with accuracy on the reconnaissance task of 71% and 29% for "D" and "O" respectively.

The pilots attended most to time, as indicated by self-evaluation accuracy. Overall self-evaluation accuracy was low for horizontal position, where all performance was "O," and errors were as large as 100 times the desired criterion. This was despite the fact that on the bob-up performance is easily judged. By contrast pilots had to infer reconnaissance performance, since they had no way to know how many targets they had failed to see. Yet the self-evaluation of reconnaissance performance was strongly related to actual performance, and the pilots were better able to judge good performance. The pilots simply neglected aircraft control in the bob-up task, and so they failed to detect even very large flying performance errors. This is an unusual finding and may reflect their priorities in performing the task and the unrealistic nature of the task. The pattern of self-evaluation deficiency supports this conclusion. Time, which can be perceived with little attention, was most accurately evaluated. Altitude error, which was strongly cued by the ground and the hover boards in front of the aircraft, was judged less well. Horizontal position required the pilot to attend to the forward hover boards and the walls located well off the right nose. This required both head movement and mental effort. The pilots were unable to integrate this activity with their other tasks. While this bob-up task left much to be desired for evaluating alerting display formats, the self-evaluations were effective in allowing us to measure the pilots' situation awareness and to relate that to the patterns of performance.

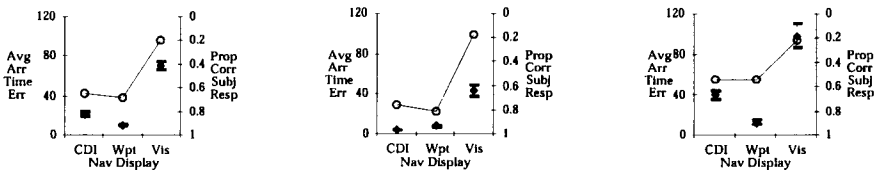
Table 4. Performance and Pilot Self-Evaluation on the Bob-up Task. Proportions Sum to less than One because of Missing Data.

Performance Factor	Proportion of Trials	Proportion of Self-Evals	Proportion of Correct Self-Evals
Time in High Hover	D/ A/ O 0.45/0.17/0.38	D/ A/ O 0.49/0.26/0.25	D/ A/ O/ DAO Combined 0.65/0.29/0.92 (0.62)
Max Z Error in High Hover	D/ A/ O 0.15/0.26/0.55	D/ A/ O 0.34/0.45/0.23	D/ A/ O/ DAO Combined 0.28/0.26/1.00 (0.50)
Max Horizontal Position Error	D/ A/ O 0.00/0.00/0.95	D/ A/ O 0.23/0.42/0.35	D/ A/ O/ DAO Combined 0.00/0.00/1.00 (0.35)

Navigation Situation Awareness

The navigation task was the only task on which the pilot's situation awareness and self-evaluation were subject to direct manipulation. This was done through the design and content of the display format on the HMD. Two aspects of the display format could affect self-evaluation and situation awareness. These were the navigation symbology (i.e., the CDI or Waypoint) which told the pilot where he was relative to the planned course, and the arrival time clock. The former allowed him to monitor and control his course-following performance, while the latter simply allowed him tomonitor his arrival time error. Both display formats did improve performance significantly over a visual navigation condition in which the pilot had only a static map (see Figure 5).

The HMD navigation display formats had a pronounced effect on the pilot's self-evaluations. The navigation display formats improved the accuracy of the pilot's self-evaluations ($F(2, 27) = 14.31$ $p < 0.05$, SAS, 1992). Self-evaluations using the two navigation display formats were significantly different from those in the visual navigation condition by a Student Neuman-Keuls test, but were not different from each other. Based on inspection of the data in Figure 5, we expected possible effects of navigation segment and of display format by segment interaction. The segment effect failed to reach significance, however ($F1, 27) = 3.1$, $p > 0.08$). The data were insufficient to test for interactions. The segment effect might have shed light on why the navigation display formats supported more accurate self-evaluations. The second reconnaissance task was much more difficult than the first, leading to worse performance. This increased reconnaissance difficulty led to poorer arrival time performance because the pilots spent more time at reconnaissance and strayed further from the course. As they spent more time on reconnaissance, the pilots became disoriented and less situationally aware, which might have been reflected in less accurate self-evaluations. So even though the performance feedback was equal on both segments, the self-evaluations would have been less accurate on the more difficult second segment. The trend in the data supports this option more than the notion that pilots simply read their performance from the arrival time clock, but the data were statistically marginal.



a. Segments 1 & 2 b. Segment 1 c. Segment 2

Open Symbols: Proportion Correct Self-Evaluations; inverted scale.
Filled Symbols: Mean Arrival Time Error and 95% Confidence Interval of Mean.

Figure 5. Performance Self-Evaluations and Actual Arrival Time Performance on Two Navigation Segments Combined (a) and on Segments 1 (b) and 2 (c) Individually.

DISCUSSION

Performance self-evaluations offer an appealing approach to the measurement of situation awareness. They are easily gathered at the end of task performance, and they require no special training on the part of the operators. The question is whether self-evaluations truly reflect situation awareness or simply the operator's ability to note and remember the required information. The answer from this research is that performance self-evaluations do reflect operator's situation awareness. The interpretation of self-evaluations, however, can be complicated.

We saw that in the case of the reconnaissance self-evaluation, pilots were able to make self-evaluations that were highly correlated with actual performance. The validity of the self-evaluation as a measure of situation awareness is shown by the fact that the pilots could accurately infer performance quality without direct feedback. Its utility lies in showing that poorer reconnaissance performance was linked to a lesser ability on the part of the pilot to evaluate the thoroughness of the search.

On the bob-up, the pilots found the task ensemble nearly impossible to perform. In this case the self-evaluations were not strongly related to performance, but there was a strong indication from the self-evaluations that the poor performance arose from the pilots' not attending to the flying task.

The self-evaluations allowed us to examine the contribution of novel information displays to pilots' situation awareness in the navigation task. The novel displays presented course deviation and arrival time performance information. The situation awareness measure distinguished the novel displays from a baseline display, but it was insensitive to differences between novel displays.

When we examined the effect of difficulty of a reconnaissance task on navigation performance and self-evaluation, we saw that both poorer performance and poorer situation awareness resulted from the more difficult reconnaissance task. There was an indication that the pilots' self-evaluations were based more on the information in the navigation display than on the arrival time clock.

We examined the performance self-evaluation metric for situation awareness in three very different tasks, reconnaissance, bob-up, and cross-country navigation. The self-evaluation proved to be a simple, effective measure of situation awareness when analyzed with relevant performance data.

REFERENCES

Aeronautical Design Standard 33D, Handling Qualities Requirements for Military Rotorcraft, U. S. Army Aviation and Troop Command, St. Louis, MO, 1994.

Beyer, W. H. (ed.) Chemical Rubber Co. Handbook of Tables for Probability and Statistics, The Chemical Rubber Co., Cleveland, 1966.

De Maio, J. and Hart, S. G., Evaluation of Helmet Display Alert Format and Information Content, (in preparation.

Endsley, M. R. "Toward a Theory of Situation Awareness in Dynamic Systems," Human Factors, 1995a, 37(1), 32 – 64.

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

Eubanks, J. L. and Killeen, P. R. "An Application of Signal Detection Theory to Air Combat Training," Human Factors, 1983, 25, 449 – 456.

Fracker, M. L. Measures of Situation Awareness: an Experimental Evaluation, AL-TR-191-0127, Armstrong Laboratory, Wright-Patterson AFB, Ohio, 1991.

Hays, W. L. Statistics, Holt, Rinehart, and Winston, New York, 1963.

Long, G. M. and Waag, W. L. "Limitations on the Practical Applicability of d' and β Measures," 1981, Human Factors, 23, 285 – 290,.

McMillan, G. R., Bushman, J., and Judge. C. L. A. "Evaluating Pilot Situational Awareness in an Operational Environment," Situation Awareness: Limitations and Enhancement in the Aviation Environment, 1996, AGARD-CP-575 (K1-1 – K1-6), Nueilly-Sur-Seine, France.

SAS/STAT User's Guide, Version 6 (2), SAS Institute Inc, Cary, NC, 1992.