2011

Assessing Crewworkload on an Instrument Meteorological Approach into a Nonradar Airport

Marilyn French-St. George Ph.D.

ASSESSING CREWWORKLOAD ON AN INSTRUMENT METEOROLOGICAL APPROACH INTO A NON-RADAR AIRPORT

Marilyn French-St. George, Ph.D.

Transportation Safety Board of Canada

Gatineau, Canada

European Air Traffic Management Program (CARE, 2003) recommendations for a 3- phased approach to workload assessment provided Transportation Safety Board investigators insight into how operating conditions for approaches into a non-radar airport under instrument meteorological conditions impact crew workload.
It was possible to develop and use secondary task questions for three of four cognitive task domains. Qualitative assessment of verbal responses illustrated how crews use verbal information to support mental models. A trend towards longer response times for the cognitively more demanding questions supported the hypothesis that maintaining situational awareness of flight status within the approach sequence is cognitively more demanding than monitoring flight control indicators. Changes in heart rate variability could be linked to changes in task demands. NASA Task Load Index data provided quantitative and qualitative indicators of overall workload and demonstrated that high workload conditions can be triggered by a variety of operational conditions.

In the course of an occurrence investigation, A09W0037 (2009), Transportation Safety Board investigators were interested in quantifying crew workload on an approach to a non-radar airport under instrument meteorological conditions. The approach included an unusual hold configuration. The CRJ705 aircraft was hand flown from glide-slope intercept using a Head-up Guidance System (HGS). Investigators were interested in gaining insight into the relative crew workload in this condition compared to a standard ILS + autopilot approach into the same airport.

In 2003, the European Organization for the Safety of Air Navigation recommended an integrated approach to the assessment of crew workload including performance-based measures, subjective ratings and physiological arousal measures (CARE-Integra-TRS-130-02-WP2, 2003). The rational for developing a 3-step approach to workload assessment acknowledged that workload cannot be implied from task analysis alone as the cognitive resources applied to a task will differ markedly between experienced and novice operators. Similarly, measures of individual effort or arousal in response to task load may also not uniquely reflect cognitive workload as the operator may not increase effort level to meet operational demands. Finally, accuracy of primary task performance will not provide evidence of the cognitive reserve available to handle unexpected events.

Methodology

Simulation trials were conducted using the CRJ simulator at the CAE training facility in Toronto, Canada. Two volunteer crews (matched in age and experience with the occurrence crew) were instructed that they would be flying simulated approach and landings into Whitehorse airport under Instrument Meteorological Conditions. They were given time to review the Jeppesen plates prior to entering the simulator. The crews were assigned one of two simulation trial sequences: HGS followed by Auto-pilot +ILS or Auto-pilot +ILS followed by HGS.

Based on CARE, 2003 recommendations three sets of measurements were taken:

1. Secondary Task Performance
   During each trial, a series of probe questions were presented to each Captain and First Officer. The probes were designed to challenge the crew's cognitive awareness. Each probe challenged one of three cognitive tasks domains based on the categorization strategy described by Anding (2008). Domain 1 probes challenged the crew's awareness of operation conditions that are considered to be attributes of tasks (primary tasks) within the current focus of attention. Domain 2 probes challenged the crew's awareness of events that are within the crew's current operational condition. Domain 3 probes challenged the crew's current situational awareness of

the flight within the greater context of the approach (table 1). The time interval between the end of the question and the start of the response defined the response time.

2. Heart Rate Variability
   The low frequency spectral power of all NN intervals between 0.04 and 0.15 Hz (LF) is the recommended Heart Rate Variability measure. Sampling was performed over successive 5 minute intervals via portable Holter monitors fitted by a trained technician.
3. NASA Task Load Index (TLX)
   The TLX was administered in a pencil and paper format in the cockpit immediately following each trial (http://humansystems.arc.nasa.gov/groups/TLX/). The task context for the TLX ratings was specified as after intercepting the glide slope. The TLX were estimated by each crew member for both primary and secondary tasks. PF: primary task: Establish and maintain stabilized flight, PF secondary task: Error trap miscommunications and missed communications. PNF primary task: Perform all radio and intercom tasks. PNF secondary task: Monitor aircraft performance and provide feedback to PF re departures from stabilized flight

Table 1.
Sample Cognitive Domain questions.

Secondary Task Probes		**Cognitive Domain**
PF	**PNF**	
Captain, what is your current altitude?	First Officer, what is the current wind direction and speed at this altitude?	1
Captain what was the FO's last call back to ATC	First Officer, what is your current altitude and estimated time to missed approach?	2
Captain what is the traffic ahead of you	First Officer, what is the timing from Robinson inbound to the missed approach point?	3

Note. Domain 1 challenges crew's awareness of operation conditions that are considered to be attributes of tasks (primary tasks) within the current focus of attention. Domain 2 probes challenged the crew's awareness of events that are within the crew's current operational condition. Domain 3 probes challenged the crew's current situational awareness of the flight within the greater context of the approach.

In addition to these three measures, a fourth subjective assessment of flight deck mutual awareness was administered after each simulation trial (Figure 1).

PF

During the period from interception of the glide path to touchdown, how aware was the PNF of your operational conditions?

1	**2**	**3**	**4**	**5**	**6**	**7**
Completely Unaware		Somewhat Unaware		Somewhat Aware		Completely Aware

PF

During the period from interception of the glide path to touchdown, how aware were you of the PNF's operational conditions?

1	**2**	**3**	**4**	**5**	**6**	**7**
Completely Unaware		Somewhat Unaware		Somewhat Aware		Completely Aware

Figure 1. Mutual Awareness Assessment scales for the PF. The upper scale allows the PF to estimate how aware he thought the PNF was of the PF's operational conditions. The Lower Scale allows the PF to estimate his awareness of the PNFs operational conditions. A similar set of scales were presented to the PNF.

Results

Secondary Task Analysis.

Figure 2 illustrates the average response times to cognitive domain questions by the PFs and PNFs in both HGS and ILS conditions. Responses to Cognitive Domain level 3 questions were significantly longer than level 1 questions (p=0.00125). These data suggest that we were successful in designing probe questions that challenged different cognitive demand levels. There was no significant difference between responses in the HGS compared to the ILS operational condition indicating that both conditions provided similar cognitive challenges.

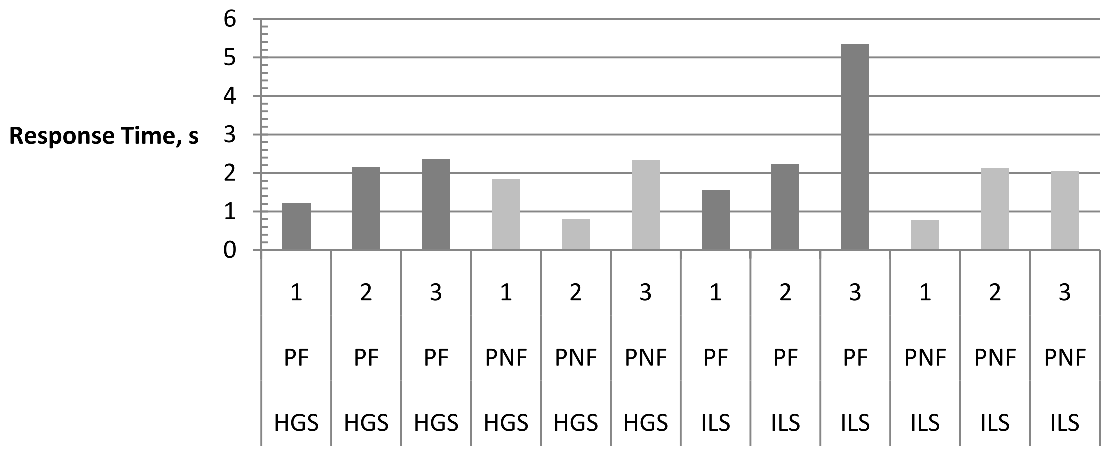

Figure 2. Average response time in seconds to cognitive domain questions (levels 1, 2, and 3) by PFs and PNFs in HGS and ILS operational conditions

Content analysis (Table 2) of the responses revealed similar recall deviations to level 2 and 3 cognitive domain questions in HGS and ILS conditions. The PFs typically did not provide all components of tower communications and confused the sequencing of Tower and Center communications. PNFs delayed call-back request by Center and miscalculated missed approach timing. Whitehorse timing for missed approach should be referenced to the Robinson NDB which is 17 miles from the airport.

Table 2
Sample responses from PFs in response to level 2 and 3 cognitive domain questions

Probe Question	**Response**	**Interpretation**
Captain please recall as many details as you can about the most recent ATC call	C1 ah The call was to contact Tower... cleared for the approach #2 contact Tower and second radio call him back at 9000 feet.	The very last communication from ATC was that a CRFI report would be available when contact with tower is made. The response given was for the second last call which contained the most recent clearance and instructions. Captain recalled 4/7 items
Captain what was the FO's last call back to ATC	Ah… He acknowledged the maintain this heading	FO also read back the altitude of 11000 and that the hold clearance would be reissued. Captain remembered approximately 1/3 of what FO read back.
Captain what does the tower know about the status of your approach	C1 ah knows that we are 9000ft and does not know that we have crossed the final approach fix yet	The tower knows that they are on the ILS31L and are to call 10 miles back. His response about 9000 feet is related to the centre controller.
	C2 ahhh we called him 10 miles and he needs an extra call at five and we are on the ILS	Captain did not recall that FO made call at 6 miles out

Heart Rate variability

Figure 3 illustrates the Low Frequency (LF) heart rate variability measures for the first and second crew for both HGS and ILS trials. As mental workload goes up, the LF heart rate variability measure goes down.

The first trial for Crew 1 was curtailed due to a SIM malfunction that induced an abrupt missed approach response from the crew at 7:40 pm. A sharp downward dip occurs in the heart rate data at approximately 7:40 pm for both the captain and the first officer indicating that work load increased quickly just before the trial was stopped. Additional dips can be seen in the Captain's trace between 7:05-7:10 and 7:25-7:30. These time intervals correspond to Edmonton Center initially updating the hold sequence and the communications transfer between Edmonton Center and Tower respectively.

The second trial (ILS) starts with both pilots indicating relatively low mental workload compared to the first trial. This is not unexpected as the crew is now more familiar with what will be expected of the trial. However, the captain's HF curve dips sharply between 8:20 and 8:25 pm which corresponds to the time when the crew realized that they intercepted the localizer above the glide path. This required manipulating the Flight Management System to increase the rate of descent to intercept the glide path from above.

For the second crew, the upward trending of the heart rate variability measures appears to indicate an easing of the workload for the second trial. In this trial, it would appear that the Captain's workload is high and is maintained throughout the trials. There are two possibilities to explain the relatively flat heart rate variability data demonstrated by the second captain. Firstly, this captain provided significantly more verbal expression of his thought strategy which he shared with his first officer. It was clear that he was thinking ahead out loud and maintaining awareness of the FO's understanding of the flight status. Secondly, the act of talking itself can serve to disrupt the heart rate variability measure.
The apparent lowering of workload for the first officer between 9:05-9:10 and 9:55-10:00 corresponds to being in revised hold patterns at Robinson and ELTAG respectively.

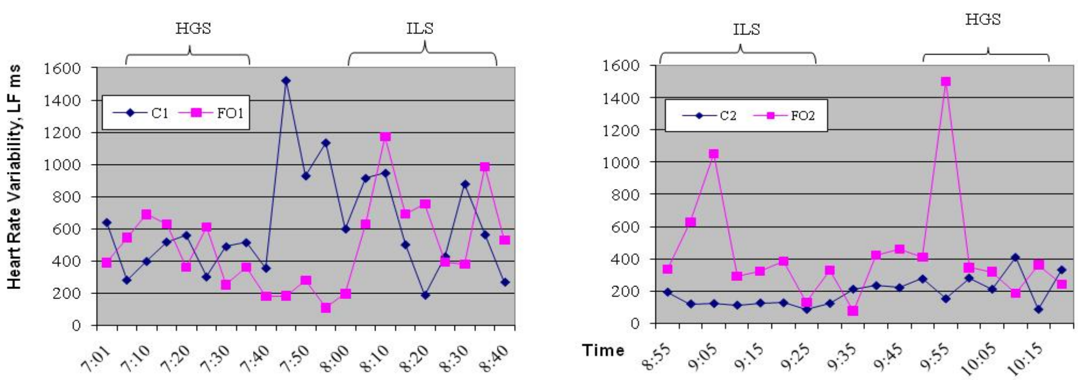

Figure 3. Heart Rate variability measures for C1/FO1 and C2/FO2 as a function of time

NASA Task Load Index measures

Tables 5 and 6 tabulate the overall TLX ratings for the first and second crews respectively. The first Crew generally assessed the task load to be somewhat higher in the ILS condition than the HGS condition. The captain volunteered that the increased workload was largely attributable to the fact that he took Whitehorse clearance above glide slope and had to correct for it on landing.

The second crew rated the overall Task Load slightly higher under the HGS condition compared to the ILS condition. These findings are consistent with the heart-rate variability measures where the second crew captain clearly showed reduced heart rate variability consistent with a higher mental task load.

Table 3

Overall TLX Ratings in percent for Crew 1 and 2

Overall TLX Rating, %		
	HGS	ILS
FO1P	66	73
FO1S	49	51
C1P	57	72
C1S	65	79

Overall TLX Rating, %		
	HGS	ILS
FO2P	75	69
FO2S	74	68
C2P	72	64
C2S	70	67

The second crew did demonstrate a small, systematic shift towards higher workload ratings in the HGS mode. Their HGS scores were at or above 70 which is considered to be the threshold for high workload (Hancock, 2009). According to Hancock there are no guidelines as to how long high workloads can or should be sustained. While this crew did show a small effect for the HGS mode, the first crew did not. Their ratings were far more influenced by the operational demands of intercepting the glide slope at an altitude somewhat higher than optimum.

Mutual Awareness Ratings

The final assessment component of mutual awareness indicated that crews appear to have sufficient self awareness of their ability to monitor their own operational conditions. In moments of high workload there appears to be a significant risk that crews will overestimation the ability of the other pilot to maintain situational awareness levels.

Crew	Condition	My awareness of his operation conditions	His awareness of my operational conditions
C1	HGS	6	7
FO1	HGS	6	6
C1	ILS	3	6
FO1	ILS	5	6
C2	HGS	5	6
FO2	HGS	6	6
C2	ILS	6	6
FO2	ILS	7	6

Overestimation

***Figure** 4* Mutual awareness ratings for both crews in each condition

Observations

The current data suggest that use of specific cockpit technologies per se is unlikely to be intrinsically associated with high workload. Rather, there are likely to be significant inter crew, and possibly inter-pilot differences in perceived work load based on experience and comfort level with the technology. Furthermore, operational decisions such as descending to meet the glide path after intercepting the localizer produce similarly high work load conditions as unfamiliarity with particular cockpit technologies. Given that the occurrence captain volunteered that he was “locked-on” to the Heads-Up Guidance System display, and the occurrence flight was the First Officers first live HGS approach, it is possible that they were both approaching a performance-based maximum workload for the duration of the flight after intercepting the glide slope. Finally, mutual awareness rating data indicate that when high workload conditions arise, crew members may become aware of their colleague’s more channeled attention but may not understand the impact that it has on mutual awareness. As a consequence there appear to be no strategies to facilitate the restoration of divided attention behaviors necessary to maintain optimum situational awareness.

References

A09W0037 (2009) Air Investigation Report: Risk of Collision between Air Canada Jazz Bombardier CL-600-2D15, C-FDJZ and Airport Maintenance Vehicles (Snow Sweepers) Whitehorse International Airport, Yukon 06 March 09.

Anding, P.N. (2008) A generalized model of cognitive task categorization, Retrieved from http://www.kmice.uum.edu.my/kmice08/listofPapers.html

CARE-Integra-TRS-130-02-WP2 (2003). Review of Workload measurement, Analysis and Interpretation Methods, retrieved from http://www.eurocontrol.int/integra/gallery/content/public/documents/expt_def_wp2_final.pdf

Hancock, Peter (2009). Personal Communication. Provost's Distinguished Research Professor, Applied experimental Human Factors, University of Florida