2007

The Use of Physiological and Subjective Measurements on the Evaluation of Cockpit Workload: A Comparison Between Simulated and Real Flight

Carlos Gomes de Oliveira

THE USE OF PHYSIOLOGICAL AND SUBJECTIVE MEASUREMENTS ON THE EVALUATION OF COCKPIT WORKLOAD: A COMPARISON BETWEEN SIMULATED AND REAL FLIGHT

Carlos Gomes de Oliveira
Civil Aviation Ergonomics Studies Management – Brazilian Agency of Civil Aviation
Rio de Janeiro – RJ – Brazil
Selma Leal de Oliveira Ribeiro
Civil Aviation Ergonomics Studies Management – Brazilian Agency of Civil Aviation
Rio de Janeiro – RJ – Brazil

The interest in determining pilot's workload has increased, mainly when considering the human factors certification issues of new aircraft, where the high insertion of automation devices has been proposed to increase the mental workload (MW). The use of subjective and physiological measurements for pilot workload evaluation has been previously investigated in a simulator, and further tested during real flights, but no comparison has been performed between the results of such different situations. This study compared the data of subjective and physiological measurements obtained on the evaluation of MW in simulators, with those measured during real flights. Seven pilots performed six different tasks in the simulator, and two pilots flew six flights. These flights were performed aiming at the certification process of a new aircraft regarding human factors, and, as certification requirements, they were conducted with the aircraft set under abnormal condition. In the simulator all pilots performed the same tasks, while assuming both, the pilot flying (PF) and pilot monitoring (PM) positions. During the flights the pilots exchanged the PF and PM position depending on flight scheduled. The pilots' ECG was registered in a computer, and the heart rate variability (HRV) processed for each task of simulator, and take-off and landing of flights. From the power spectral density function of HRV it was determined the total power of low frequency (LF) and high frequency (HF). The HRV was analyzed as the ratio of the LF/HF obtained during the evaluated task (tasks of simulator; take-off and landing of flights) and the LF/HF values obtained during a rest test, which was registered either prior to the flight or to the section of simulator. The NASA-TLX Scale was applied just after the task was finished. The results showed the NASA-TLX score of PM in the simulator to give higher values than PF, but this was not statistically significant. The same tendency was observed on HRV, but most of the tasks reached statistical significance ($p<0.05$). Conversely, during the flights PF showed higher values of NASA-TLX score and HRV than PM, although no statistical test was applied in such data due to the number of pilots. Surprisingly, the NASA-TLX showed higher values in simulator than during the flights. The data from simulator tasks and flights suggests the mental demand of NASA-TLX to be correlated to HRV, but HRV appears to be more sensitive to different intensities of mental workload. The differences observed when comparing simulator and flights were more likely due to the nature of the tasks, level of automation, or even the workload perceived by pilots. It is concluded that the use of both subjective and HRV measurements give a potential tool in the evaluation of pilot MW, and could further be used to quantify the workload objectively, defining acceptable ranges of MW that pilots are subjected to. The assessment could be performed in simulator since this machine presented results very compatibles with those found in real flights.

Introduction

The complexity of the task performed by pilots in aviation has changed due to the improvement of automation in the cockpit. The modifications performed so far appears to look at a better distribution of workload during the flight (WISE & WISE, 2000; TATTERSALL, 2000). It is necessary to maintain a balance between the demand of a given task and the capacity of the operator according to the objectives of the tasks, which include those tasks applied to evaluate items related to the certification of new aircraft regarding human factors issues.

One important concern that has been attempted is how to properly measure the workload of pilots in cockpits, and how to establish its minimum and/or maximum level allowed, specially regarded the mental workload, whereas the literature shows a consistent search on this subject, and subjective and physiological methods of measurements have been employed (BACKS, 1995; MIYAKE, 2001).

Ribeiro & de Oliveira (2003) proposed a method for evaluation mental, physical and overall workload in pilots, which was firstly experimented in simulated flights, and was further applied in real flights, performed during a certification process of a new developed aircraft (de OLIVEIRA & RIBEIRO, 2005). One aspect that one might consider when using a simulator for workload assessment is that although its use plays an important rule on pilot training, investigations on how or even if simulator could be appropriate to evaluate pilot workload has not extensively been accomplished and few studies have focused on this subject.

The present study used the same data of Ribeiro & de Oliveira (2003), and de Oliveira & Ribeiro (2005) to compare the mental workload (MW) of tasks, performed in a simulator, with that of real flights, performed during the last phase of the certification process of an aircraft, aiming at identify differences and similarities on the results obtained by the methods, previously reported as potential tool for workload evaluation.

Methods

The whole procedures of data collection were described previously (RIBEIRO & de OLIVEIRA, 2003 and de OLIVEIRA & RIBEIRO, 2005), and thus they will be summarized as following.

Simulator Evaluation. Seven pilots with good experience on the simulator of the EMB120 cockpit participated in this study. The simulator provides virtual situation conditions as real as a true flight in an EMB120 aircraft. Each subject performed the tasks as described bellow, assuming the pilot flying (PF) and pilot monitoring (PM) positions.

Task 1 (T1) – Normal take-off. It finished when the aircraft reaches 400ft with zero flap.

Task 2 (T2) – Take-off with one engine failure (right or left) at V1. It finished at 1000ft.

Task 3 (T3) – ILS approach with single engine and Flight Director.

Task 4 (T4) – ILS approach with single engine and no Flight Director.

Task 5 – ILS approach with two engines and Flight Director.

Task 6 - ILS approach with two engines and no Flight Director.

The data was collected in eight sections, in which all six tasks were developed by each pilot, one as PF and one as PM. A given pilot who assumed the PF position did not participate in the same section as PM with the same pilot who assumed the PM position in such section.

Real Flight Evaluation. This part of the study was conducted during the certification process of a new developed aircraft made in Brazil. Only two high experienced pilots were monitored. They alternated the position of PF and PM, but not in the same flight. Six flights were monitored during three consecutive days, and from all flights two phases were evaluated: take-off (began when the engine one was switched on and finished when the aircraft reached 15,000ft), and landing (began at 10,000ft and ended when all engines were turned off). The routes and abnormal situations that occurred during the flight, which included the absence of electric, hydraulic and other automated systems, were determined by the team formed by the certification authority, and the manufacturer technical staff.

Instruments of evaluation

As there is a considerable amount of variables in both studies from which this work took the data (RIBEIRO & de OLIVEIRA, 2003; de OLIVEIRA & RIBEIRO, 2005), only mental workload was evaluated through physiological and subjective techniques.

Physiological Evaluation. The power spectral analysis of Heart Rate Variability (HRV) signal is a sensitive index to the autonomic nervous system activities. In this signal two main components have been identified, the low frequency (LF) at 0.03-0.15Hz, reflecting both sympathetic and parasympathetic activity, and high frequency (HF) at 0.15-0.4Hz, which reflect the parasympathetic tone of the sinusoidal respiratory arrhythmia. LF/HF ratio has been proposed as an index that reflects the balance of the autonomic nervous activity (TASK FORCE of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996). Moreover, previous studies have revealed an increase of this index when individual is subject mental effort (SATO *et al*., 1998; KAMADA *et al*. 1992). Thus, in the present research the physiological evaluation was performed through the measurement of the LF/HR of the HRV signal. The electrocardiogram (ECG) signal was captured and simultaneously digitally recorded, after sampled at 1000Hz. A specific program to detect the R-waves of the ECG signal and construct the RR intervals was developed in Matlab 5.02c (Mathworks). The time series formed by the RR intervals were thus interpolated so as the sample rate of the respective HRV signal was 2Hz and the total power of LF and HF bands of its power spectral density were determined. Prior to each task of simulator and flights, the ECG of the pilots were registered during a rest period of 4 minutes, and the LF and HF estimated. The LF/HF of the tasks was thus divided by the respective rest test to normalize the data and improve the inter-subject variability. This ratio thus was taken as the HRV variable evaluated.

Subjective Evaluation. To evaluate mental workload the subjective techniques are more often applied, and can be considered as indices of global sensitivity to the workload (WIERWILLE & EGGEMEIER, 1993). The subjective technique used in this study was the Task Load Index Scale - NASA – TLX (HART & STAVELAND, 1988), which has been showed to be consistent in many studies with different levels of demand (HARRIS *et al*., 1995; HANCOCK *et al*., 1995). Only the mental demand of such scale was assessed, using a score ranging between 0 and 20, obtained without the use of weighting procedure, as suggested by Nygren (1991). The NASA-TLX was applied after the end of each task of simulator, and after take-off and landing of the flights evaluated.

As many statistical procedures was previously reported for the simulator tasks (RIBEIRO & de OLIVEIRA, 2003), in the present study only comparison of PF against PM was performed through *Wilcoxon* matched pair test. Since only two pilots participated on the real flight situations no statistical procedure was applied in such data.

Results

The NASA-TLX showed a tendency of higher scores when the pilot assumed the PM position in almost all tasks performed in the simulator (Figure 1), but none reached statistical significance ($p>0.05$). This lack of statistical significance was more likely due to the high inter-subject variability found (Figure 2), although even when considering all tasks together no statistical significance was reached.

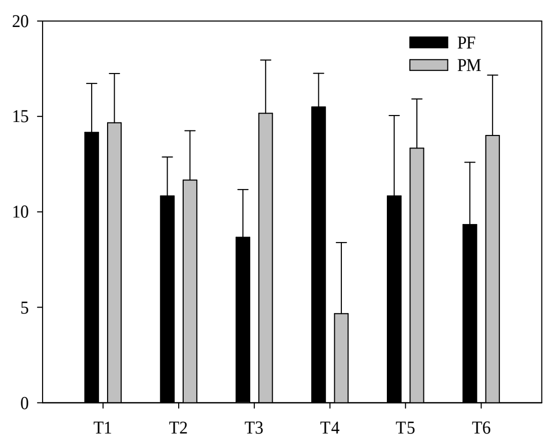

Figure 1. Results of the mean (bar) and standard deviation (line on top of the bar) of NASA-TLX mental score of all tasks performed in the simulator

The HRV showed the same tendency, which is, higher values for the pilots while PM than PF. However, differently of the NASA-TLX scores, such variable reached statistical significance in almost all tasks (Figure 3). Despite the considerable high inter-subject variability also observed in this variable, again, differently of NASA-TLX, when considering all tasks, the statistical significance was high ($p=0.004$).

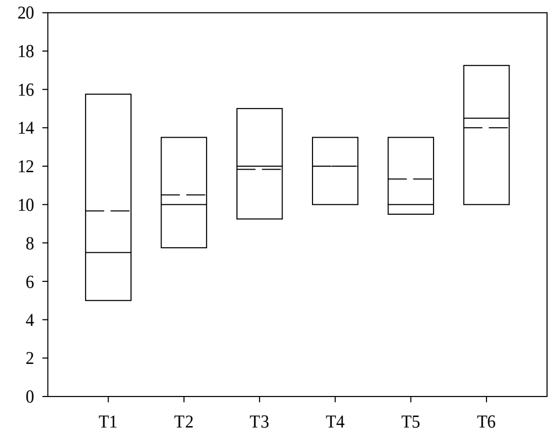

Figure 2. Box Plot of the mental score of NASA-TLX of all tasks performed in the simulator showing the 5% and 95% confidence interval, the mean (dashed) and median (filled) line

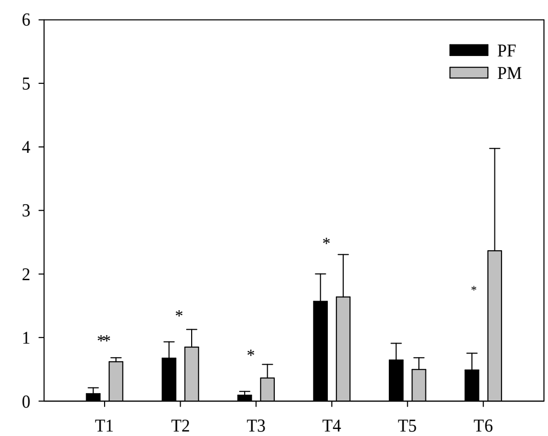

Figure 3. Results of the mean (bar) and standard deviation (line on top of the bar) of the normalized HRV of all tasks performed in the simulator. Significantly higher value was accounted in PM than in PF for $p<0.05$ (*) and $p<0.01$ (**)

In order to better show the results of real flights they are numbered from 1 to 6 (some data from the flight number 6 was missed). When one of the two pilots (P1 and P2) assumes the PF position in the forward direction of a route this is labeled as A. In the same flight, the other one pilot assumes the PF position in the backward direction of the same route, and this is labeled as B.

The NASA-TLX score showed that, as PF, P1 presented higher MW during most landings than take-offs (Figure 4), and as PM, the MW alternated in intensity during take-off and landing, and this did not depend on type of flight (Figure 5). The pilot P2 presented a different behavior, since when assumed the PF position showed low MW, and had almost the same values during take-off and landing (Figure 5). This was also true for this pilot while in the PM position (Figure 4). These results indicate that different abnormal situations had no effect in the mental workload perceived by this pilot.

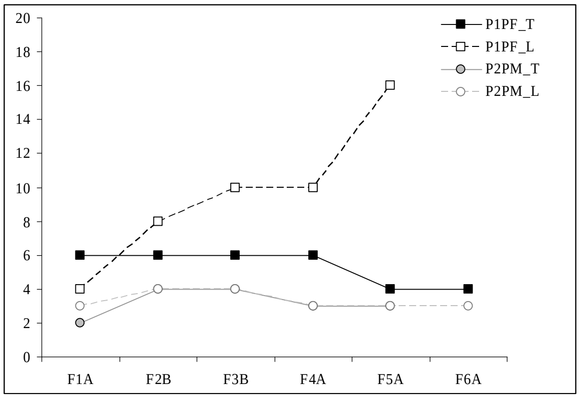

Figure 4. Results of NASA-TLX score according to flight (abscissa axis), pilot (P1 and P2), position (PF, PM), and take-off (T) and landing (L)

The physiological measurement corroborated the results of many findings of the NASA-TLX. Pilot P1, as PF showed higher values of HRV during landing than take-off (Figure 6), suggesting higher mental workload during landing. On the other hand, as PM, HRV of P1 did not show clear pattern (Figure 7). The HRV also tended to be higher during landings than take-offs in pilot P2, as assuming the PF position (Figure 7). When in the PM position, again no pattern was observed (Figure 6), and it should be pointed out that in one flight the hear rate of this pilot was lower than that presented in the rest test, performed prior to the flight.

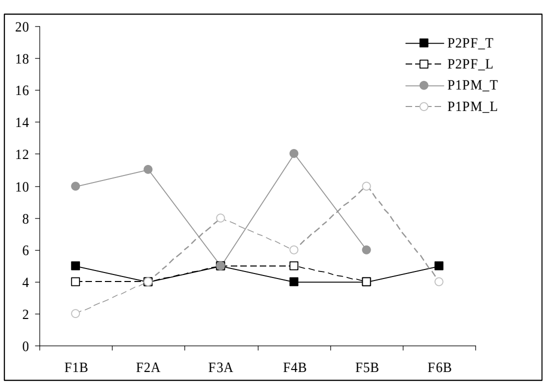

Figure 5. Results of NASA-TLS score according to flight

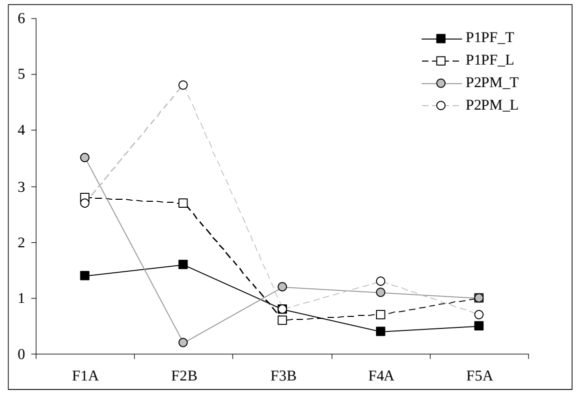

Figure 6. Results of HRV according to flight

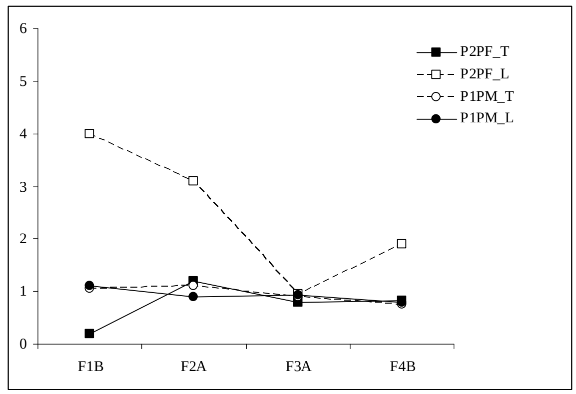

Figure 7. Results of HRV according to flight

Although statistical analysis is not applicable, one can observe that the mean score of NASA-TLX tended to be lower during real flights than during simulator tasks, even with the flights set in abnormal situations, and that the PF was more likely to have high MW than PM, whereas the contrary was observed in the simulator (Figure 8). It is important to stress that even though in the simulator no statistical significance was observed for NASA-TLS, the HRV results reached this significance.

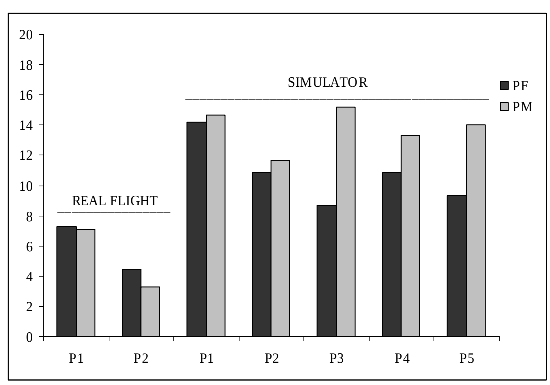

Figure 8. Results of the score of NASA-TLX of real flights and of the simulator (the five pilots who presented lowest values)

Discussion

One important concern that has been raised is how to have a consistent and appropriate tool for measure pilot workload during flight, mainly regarded to certification of new aircrafts for human factor aspects. Mental or cognitive workload was proposed as difficult to be assessed (KANTOWITZ & CASPER, 1988), and although one could identify in the literature a consistent search for evaluation the mental workload by the use of subjective and physiological methods (BACKS, 1995; MIYAKE, 2001) a “standard gold” does not appear to be established.

Ribeiro & de Oliveira (2003) reported that the use of physiological associated to subjective measurements could be a good tool in such subject, and a similar approach was adopted during real flight, being the results promising for the search of such a method (De OLIVEIRA & RIBEIRO, 2005). The present study simple evaluated the main aspects that cause differences in the data obtained by those studies.

The first observation to be pointed out is the lack of statistical significance of the NASA-TLX scores given by the pilots who participated on the measurements in the simulator. As was identified, the high inter-subject variability (high variance) could be the more probable source of such findings. However, one could not forget that the nature of the subjectivity inherent to pilot perception can be the most source of this variability. On the other hand, as HRV showed a consistent and similar pattern, which is higher values when pilot assumed the PM than PF position, this reached statistical significance. Veltman & Gaillard (1998) proposed that there is more respiratory activity during rest than during mental tasks, reflecting in an increase of power of LF band. This would imply in a higher LF/HF ratio during mental workload, which was also observed by Hjortskov *et al* (2004) in a similar way, but regarded to mental stress. Sato *et al.* (1998) and Kamada *et al.* (1992) also have proposed that during mental effort this ratio tend to increase when compared to the rest. Although some results of the tasks performed in the simulator were lower than unity, as the perceived mental workload increased, so did the HRV, showing that this physiological measurement is in accordance to the perception, and the statistical significance found suggests that this probably is more sensitive to changes in mental workload than subjective assessment.

When considering the real and abnormal flights, one might speculate why mental workload on such situation was perceived as lower than in a simulator. De Oliveira & Ribeiro (2005) proposed that the automation of the aircraft under certification could explain this apparent paradox. Furthermore, the pilots who performed the flight were highly experienced as test pilots. However, when looking at the results, it should be highlighted that whereas the perceived workload by the pilots, mainly pilot P2, had low differences, considering flight and position, the HRV showed high differences, mainly regarding position and task (take-off and landing). This could lead us to interpret that the perception itself is not as sensible as it is the physiological response to the task.

The second aspect to be discussed is why PF showed higher MW than PM during the flights, and the contrary was observed in the simulator. The lower automation of simulator when compared to the aircraft (highly automated) used during the flights could be an explanation for such findings, because low automation probably requires more supervision of PM. Another aspect that should be taken into consideration is that some tasks of the simulator were developed to provide high demand of workload on both PF and PM, while the flights were performed to certificate an aircraft. However, it should be pointed out that flights performed with abnormal situations, had lower effect in terms of create mental workload than a simple tasks in a simulator. This leads one to believe that the use of simulator does not only play an important rule on pilot training, but also is able to replicate scenarios when one intends to create realistic conditions for pilot mental workload assessment. One question to be further answer is how to quantify the workload in an objective criterion, defining acceptable levels.

Conclusion

The HRV appears to be very sensible to mental workload assessment in aviation, and when used together with a subjective scale can provide orientation to those who are interested in such evaluation. The methods presented in this study should be improved looking at establish ranges of mental workload that pilots are subject to, which could be conducted in simulators, since this machine showed to have results comparable to those obtained during flights, which might be previously believed to impose high complexity tasks to the pilots.

Acknowledgments

This study was developed with the support of the Brazilian Agency of Civil Aviation through the Technical Cooperation Mission of the International Civil Aviation Organization and the technical cooperation of the Aeronautical Certification Division.

References

BACKS, R. W. (1995) Going beyond heart rate: autonomic space and cardiovascular assessment of mental workload. In: *The International Journal of Aviation Psychology,* 5 (1), 25-48.

de OLIVEIRA, C.G., and RIBEIRO, S.L.O (2005) Evaluation of workload in high complexity work place: an experiment during a real situation. In: XIIIth International Symposium on Aviation Psychology. (*Proceeding*s) Oklahoma, OK: Wright State University, 177-182.

HANCOCK, P. A., WILLIAMS, G., MANNING, C. M. (1995) Influence of task demand characteristics on workload and performance. In: *The International Journal of Aviation Psychology*, 5 (1), 63-86.

HARRIS, W. C., HANCOCK, P. A., ARTHUR, E. J., CAIRD, J. K. (1995) Performance, workload, and fatigue changes associated with automation. In: *The International Journal of Aviation Psychology*, 5 (2), 169-185.

HART, S. G. & STAVELAND, L. E. (1988) Development of NASA-TLX (task Load Index): results of empirical and theoretical research. In: HANCOCK, P. A. & MESHKATI, N. (Eds) *Human Mental Workload*. Amsterdam: Elsevier.

HJORTSKOV, N et al., (2004) The effect of mental stress on heart rate variability and blood pressure during computer work. In: *Eur J Appl Physiol*, 92(1-2), 84-89.

KAMADA T, KUMASHIRO, M; MONOU, H; INOUE, K. (1992) Power spectral analysis of heart rate variability in Type As and Type Bs during mental workload. In: *Psycosom Med*, 54, 462-470

KANTOWITZ B.H. and CASPER, P. A. (1988) Human workload in aviation. In: WIENER, E. L. & NAGEL, D. C. (Eds.) *Human factors in aviation*. San Diego, California: Academic Press:.

MIYAKE S. (2001); Multivariate workload evaluation combining physiological and subjective measures. In: *Int J Psycophysiol*. 40 (3), 233-238.

NYGREN, T. E. (1991) Psychometric properties of subjective workload measurement techniques: implications for their use in the assessment of perceived mental workload. In: *Human Factors*. vol. 33 n º 1, p.17-33.

RIBEIRO, S.L.O and de OLIVEIRA, C.G. (2003) Workload in flight cockpit: an approach for searching a methodological evaluation. In: XIIth International Symposium on Aviation Psychology. (*Proceedings*) Dayton, OH: Wright State University, 993-998.

SATO, N.; KAMADA, T.; MIYAKE, S.; ARAKATSU, J.; KUMASHIRO, M.; KUME, Y. (1998) Power spectral analysis of heart rate variability in type A females during a psychomotor task, In:*Journal of Psychomotor Research*. 45(2), 159-169.

TASK FORCE of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996) Heart Rate Variability: Standards of Measurement, Physiological Interpretation and Clinical Use. In: *Circulation* 93(5):1043-1065.

TATTERSALL, A. J. (2000) Practical guidelines for workload assessment. In: WISE, J. A. and HOPKIN, V. D. (Eds.). In: *Human Factors in Certification*. New Jersey: Lawrence Erlbaum Associates, Publishers.

VELTMAN, H. & GAILLARD, A. (1998) Physiological workload reactions to increasing levels of task difficulty, *Ergonomics*, 1998, 41(5), 565-669.

WIERWILLE, W. W., & EGGEMEIER, F. T. (1993) Recommendations for mental workload measurement in a test and evaluation environment. *Ergonomics*. 35 (2), 263-268.

WISE, M. A. & WISE, J. A. (2000) The use of the systems approaches to certify advanced aviation technologies. In: WISE, J. A. & HOPKIN, V. D. (Eds.). *Human Factors in Certification*. New Jersey: LEA.