Effects of Workload on Measures of Sustained Attention During a Flight Simulator Night Mission

Hans-Juergen Hoermann

Patrick Gontar

Andreas Haslbeck

EFFECTS OF WORKLOAD ON MEASURES OF SUSTAINED ATTENTION DURING A FLIGHT SIMULATOR NIGHT MISSION

Hans-Juergen Hoermann
Institute of Aerospace Medicine, Department of Aviation and Space Psychology
German Aerospace Center (DLR), Hamburg, Germany

Patrick Gontar and Andreas Haslbeck
Institute of Ergonomics, Technische Universität München
Munich, Germany

N=60 commercial airline pilots holding valid ATPLs flew a manual ILS approach following a weather induced missed approach during a night mission in full flight simulators. Measures of subjective fatigue, sustained attention, and the NASA Taskload Index were collected before and after the mission. In addition, sleep history data were available covering three days prior to the simulator. Both subjective and objective measures of fatigue showed significant ascent over the three hours of the experimental procedure. While sleep history data and roster information were related to both the overall level of fatigue and to reaction times, pilots who experienced a higher degree of workload during the simulator exercise showed a significant increase in subjective fatigue scores after the mission. The findings provide some evidence for lasting effects of a sleep deficit as well as for a multifactorial model of fatigue risk.

Most models of fatigue risk in aviation can be traced back to the classical two-process model of sleep regulation (Borbély, 1982), which explains sleepiness through the interaction of homeostatic (sleep pressure by time awake) and circadian influences (circadian phase). In order to achieve more accurate predictions, some fatigue risk management systems (FRMS) consider additionally sleep inertia (Åkerstedt & Folkard, 1990), task-related factors (i.e. time-zone transitions, workload, work-schedule), individual factors (i.e. life-style, chronotype) or cumulative effects (VanDongen et al., 2003). However, empirical validation data for task-related and individual factors with cognitive effectiveness within the aviation environment are rare and contradictory (Tritschler & Bond, 2010; Williamson et al., 2011).

This study aimed to explore the relationship between subjective and objective measures of fatigue with factors of workload and work scheduling. Our data were gathered before and after a simulator night mission with a sample of long- and short-haul pilots who had been awake for more than 16 hours. It was expected that individual sleep history and scheduling factors are equally related to the overall level of fatigue before and after the simulator mission. In addition to that, we analyzed whether workload as experienced by the individual pilot during the simulator mission can be identified as a moderator variable for an increase of fatigue after the mission.

Method

Experimental Procedure

This study originally aimed for measuring manual flying skills of airline pilots in a full-flight simulator (*JAR STD 1A Level D*) night mission. However, the data of flying performance itself are reported elsewhere (Haslbeck, Kirchner, Schubert, & Bengler, 2014). All participants were asked to get up as usual in the morning and not to sleep during the daytime prior to the simulator session. The procedure started at 9:30 p.m. with dinner at a restaurant located close to the simulator facility. Three pilots per night participated in the overall 20 simulator missions. Between 11 p.m. and midnight, the baseline measurement of subjective fatigue and sustained attention took place (*base*). Thereafter, the whole group of three pilots went to the simulator and a second *pre*-simulator measurement of fatigue scores was conducted for the second and the third participant (*pre*) while the first participant started the 45-minute experiment at about 12:30 a.m. After another 15 minutes, during which the simulator was reset, the next participant started the experiment at about 1:30 a.m; the third participant started at about 2:30 a.m. For all participants, a final fatigue testing was scheduled in a briefing room immediately after they finished their simulator trial between 1:30 a.m. and 3:30 a.m. (*post*). During this *post*-session participants also assessed the level of workload during the simulator mission. In this scenario, all pilots had to perform an approach scenario towards Munich (EDDM) ending with a missed approach decision due to a strong tail wind situation (Haslbeck, Eichinger, & Bengler, 2013). After the crew performed their go-around, the tower changed the runway direction. When the pilots turned to their final runway heading, we evoked a malfunction leading to a failure of the autopilot and the flight director. Consequently, the participant had to manually fly and land the aircraft (raw data ILS).

Participants

All 60 pilots participating in this study were scheduled for a full-flight simulator research experiment; participation was part of their working schedule in terms of an additional simulator event and not discretionary. That meant they were randomly selected from the crew planning department, and participation in the experiment was part of their normal duty time (Haslbeck et al., 2012). 30 long-haul captains (Airbus A340) and 30 short-haul first officers (Airbus A320), representing a wide range of experience considering the level of practice and training (listed in Table 1.) of the co-operating partner airline participated in the study. However, only data from N=57 pilots can be included in the following statistics because one simulator mission was cancelled due to technical reasons.

Table 1.

Demographical data of all participants, adapted from Haslbeck et al. (2014)

	Age		overall flight hours	
	mean	SD	mean	SD
CPTs A340 (n=27)	50.4	4.0	15,019.7	2,938.4
FOs A320 (n=30)	30.4	3.0	3,373.9	1,703.9

Measures

A number of objective and subjective measurements were administered according to the procedure described above in order to collect data on fatigue, level of attention, workload and the three-day sleep history.

NASA Raw TLX (RTLX): A German version of the NASA Task Load Index (Hart & Staveland, 1988; Hart, 2006; Pfendler, Pitrella & Wiegand, 1994) was used as a measure of subjective workload during the simulator mission. With respect to the overall demanding procedure, the RTLX was chosen which omits the comparison of the subscales. This method delivered a total score (RTLX) and six subscales reflecting three different workload factors:

- Task related:
  - *Mental demands* (TL-MD)
  - *Physical demands* (TL-PD)
  - *Temporal demands* (TL-TD)
- Behavior related:
  - *Performance* (TL-PE)
  - *Effort* (TL-EF)
- Subject related:
  - *Frustration* (TL-FR)

All RTLX scores were scaled between 0 and 100.

Psychomotor Vigilance Task (PVT): As a measure for sustained attention, a ten-minute version of the PVT (Dinges & Powell, 1985; PEBL version by Mueller, 2011) was administered on portable computers. A simple visual stimulus was presented about 7 times per minute with variable inter-stimulus intervals. Subjects were asked to press the space key as soon as they see the stimulus. Different performance scores were calculated:

- Number of lapses with reaction times > 500ms (P-LAPS)
- Mean reaction times for the 10% slowest responses (P-RT10)
- Mean reaction times for the 10% fastest responses (P-RT90)
- Overall mean reaction times (P-MRT)

Subjective Fatigue Checklist (FAT): With the FAT (Samn & Perelli, 1982) the subjective level of fatigue was assessed subsequent to the PVT. The FAT provided subjective fatigue scores between 0 (lowest) and 20 (highest). Scores above 9 are regarded as “mild fatigue”, above 12 as “moderate fatigue”, and above 16 as “severe fatigue” (Samn & Perelli, 1982, p5).

Visual Analogue Scale (VAS): A visual analogue scale (VAS) was used for a subjective alertness/sleepiness assessment. A score of 100 was labeled as “very alert” and 0 as “very sleepy”.

Sleep diaries and roster information: Subjects started writing sleep diaries three nights before their scheduled simulator mission. The recorded parameters reflected

- Sleeping time (SD-ST)
- Wakeup time (SD-WU)
- Sleep quality (SD-QU – analogue scale from 0 – "very bad" to 100 - "very good")

From this information we calculated two further scores:

- Accumulated sleep deficit (SD-DEF): 24 hours minus the time asleep during the previous three nights
- Time awake before the simulator mission in minutes (SD-TAW)

With respect to roster information we considered here two parameters:

- Number of duty days within three days before the simulator event (SD-DUT)
- Last flight over 3 or more time zones within 3 days before the simulator (SD-TZN) (Samel et al., 1995)

Reliable Change Index (RCI): In order to receive a measure of change for the fatigue scores, the Reliable Change Index (Jacobson & Truax, 1991) was calculated between the *base*-scores and the *post*-scores. The RCIs compensate change information for unreliability of measurement.

Results

All pilots had been fatigued when they came for their simulator mission. The simulator by intention was scheduled during the circadian low. At that time, the pilots had already been awake for 16 to 22.5 hours. In addition, almost half of them (28 pilots) had accumulated a sleep deficit of more than 2 hours. 14 pilots crossed more than three time zones within the past three days (SD-TZN) and another 9 pilots had just one or no off-days before the simulator (SD-DUT). The distribution scores for SD-TAW and SD-DEF are shown in figure 1 and 2.

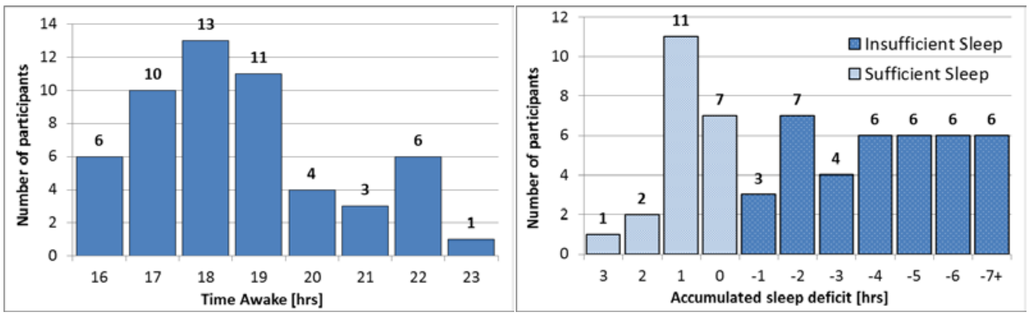

Figure 1 and 2. *Distribution of the number of hours being awake (left) and the accumulated sleep deficit over the past three days (right) before the simulator mission.*

Under these conditions of fatigue and the difficulty of the scenario, it was expected that the pilots would face a very demanding simulator mission. The mean scores of the RTLX-subscales are shown in figure 3. Subjective workload was highest with respect to M*ental Demands* and E*ffort*.

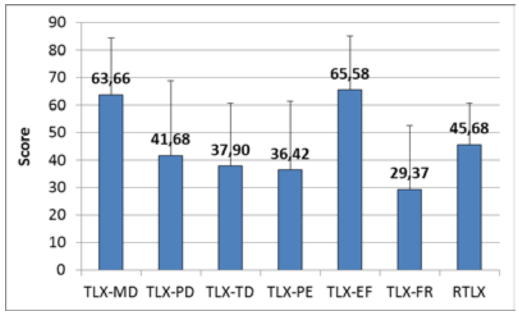

Figure 3. *Mean scores of the RTLX subscales and overall subjective workload. Error bars indicate one standard deviation.*

Paired sample T-Tests of the mean scores from the PVT and the subjective fatigue assessments showed, in several parameters, a significant ($\alpha < .05$) decrease of attention levels and an increase in fatigue. According to Cohen's d (Cohen, 1992), the effect sizes are medium for P-LAPS (d=.45), P-RT10 (d=.32), and FAT (d=.33) and small for VAS (d=.14). As illustrated in figures 4 to7, the inter-individual variances also increased from the *base*- to the *post*-measurement. FAT scores varied around 12, which means a moderate but not yet severe amount of fatigue.

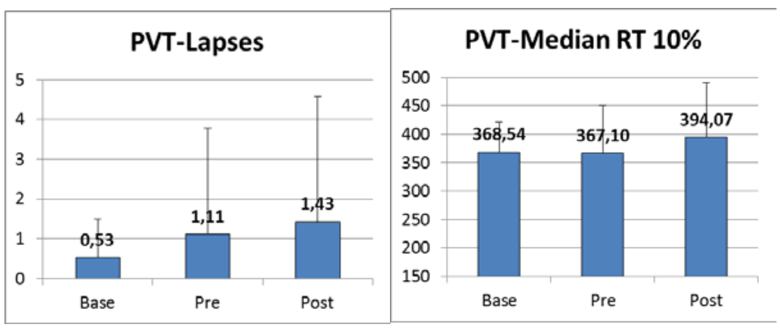

Figure 4 and 5. *Mean scores of PVT lapses (left) and the 10% slowest reaction times (right) during the experimental procedure. Error bars indicate one standard deviation.*

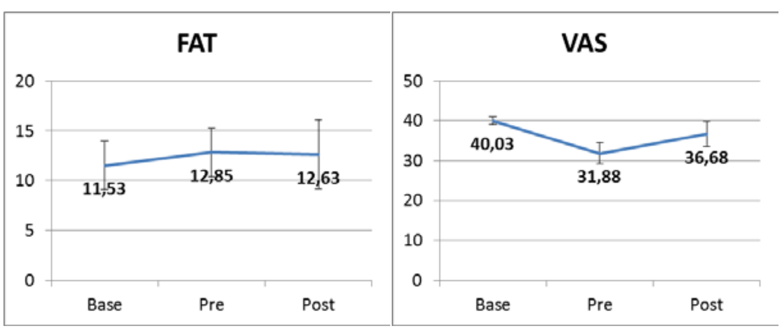

Figure 6 and 7. *Mean scores of subjective fatigue checklist scores (left) and visual analogue alertness assessment (right) during the experimental procedure. Error bars indicate +/- one standard deviation.*

Correlation analyses were conducted to explore the relationship between scores of sustained attention and workload. While the RTLX total score had no significant correlations to any PVT-scores, the task-related and subject-related RTLX-subscales showed some significant correlations ($\alpha < .05$) primarily with PVT-scores during *base*- and *pre*-measurement. Significant coefficients were for TL-PE .26 (P-LAPS$_{base}$), .23 (P-MRT$_{base}$), .41 (P-MRT$_{pre}$) and .28 (P-MRT$_{post}$). TL-EF had significant correlations of .33 (P-RT10$_{pre}$) and .32(P-MRT$_{pre}$). TL-FR showed significant correlations of .41 (P-MRT$_{base}$) and .32 (P-MRT$_{pre}$). No significant correlations with the RCI of the PVT were observed.

Looking at the correlations with subjective fatigue assessments, the relationship to workload appeared stronger. For the RTLX total score, we found significant correlations to the *post*-measurements FAT_{post} of .29 and VAS_{post} of -.24. Also, pilots who experienced a higher workload during the simulator felt increasingly more fatigued afterwards. The correlations with RCI-FAT and RCI-VAS were .33 and -.32 respectively. All significant correlations with the *post*-measurement and with the change indices are shown in table 2.

Table 2.
Significant correlations (α < .05) between subjective fatigue (level and change scores) and workload during the simulator mission

	TL-MD	TL-PD	TL-TD	TL-PE	TL-EF	TL-FR	RTLX
FAT_{post}				.35	.27	.25	.29
VAS_{post}				-.38	-.32	-.24	-.24
RCI-FAT	.25			.42	.35		.33
RCI-VAS				-.37	-.33		-.32

Further analyses of the sleep logs revealed a positive relationship (α < .05) between the number of lapses in the first PVT measurement and time awake before the simulator (r = .30). Also, the accumulated sleep deficit, which really had a wide range (see figure 2), correlated .30 with $P\text{-}LAPS_{base}$. Sleep quality one night before the simulator correlated -.30 with FAT_{base}. Sleep quality three nights prior also showed significant correlation with subjective fatigue scores during the *base*-measurement of -.28 with FAT_{base} and .26 with VAS_{base}. Sleep quality of the second-to-last night correlated .27 with RCI-VAS, which meant less decrement of alertness with higher quality of sleep.

The strongest correlations with the objective PVT-scores of sustained attention were found with the rostering information and with the time of the simulator event itself as shown in table 3.

Table 3.
Significant correlations (α < .05) between change scores of subjective fatigue and workload during the simulator mission

		Base			*Pre*			*Post*	
	P-LAPS	P-RT10	P-MRT	P-LAPS	P-RT10	P-MRT	P-LAPS	P-RT10	P-MRT
Sleep deficit	.30								
Time awake	.30								
Duty days	.28	.42	.29		.34	.41	.26	.37	.45
Time zones		.29	.25						
Sim time	.26						.24	.30	.25

The time of the simulator event as a circadian factor was identified as the strongest predictor of change between the *base*- and the *post*-measurements. The correlations are with RCI-PVT .25, with RCI-FAT .48 and with RCI-VAS -.33.

Discussion

In summary, it can be confirmed by subjective as well as objective data that the pilots were moderately fatigued when they came to their simulator mission shortly before midnight. Fatigue further increased significantly during the three to four-hour experimental procedure. The level of fatigue could be predicted systematically by rostering information (e.g. number of duty days within the three days before the simulator) and some sleep history data (e.g. time awake and sleep quality).

As illustrated by the RTLX-data, the simulator mission was above average demanding with peaks for the task load factors of *Mental Demands* and *Effort*. However, the moderating effect of workload on increased fatigue could only be demonstrated for the subjective fatigue scores (FAT and VAS, table 2). While NASA RTLX scales were significantly correlated to several PVT scores of sustained attention, there was no significant interaction between the amount of workload and the amount of attention decrements from before to after the simulator. We did not conduct explicit causal analysis here, but from the chronology of measurement it seems equally probable that workload causes attention decrements than that lack of sustained attention causes workload increments. It could be worth further investigating a common source of variance for workload and attention such as individual resources (training and basic abilities or simply the sleep history). Furthermore, individualized fitness-for-duty testing could become a promising option in this context (e.g. Elmenhorst et al., 2013).

The strongest predictor of change in levels of fatigue identified here was the time of the simulator event itself, which could illustrate the influence of circadian processes. However, with respect to our main question whether workload directly affects the levels of sustained attention, we did not find sufficient evidence. Only increases in subjective fatigue scores were significantly related to workload. An alternative explanation could be that the workload did not mount up high enough during the simulator or its effect appears with some time delay. To assess levels of individual fatigue risk in aviation, fitness-for-duty testing should complement FRMS recommendations.

Acknowledgements

Part of this work was funded by the German Federal Ministry of Economics and Technology via the Project Management Agency for Aeronautics Research within the Federal Aeronautical Research Program (LuFo IV-2). The authors acknowledge their thanks to Paul Kirchner for supporting this study and his contribution to the project.

References

Achermann, P. The two-process model of sleep regulation revisited. *Aviat Space Environ Med, 75(3*), 37–43.

Åkerstedt, T. & Folkard, S. (1990). A model of human sleepiness. In: J.A. Horne (Ed.) *Sleep '90*. Bochum: Pontenagel Press.

Borbély, A.A. (1982). A two process model of sleep regulation. *Human Neurobiology*, *1*, 195–204.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Dinges D.F., Powell J.W. (1985). Microcomputer analysis of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments & Computers*, *17*, 652–655.

Elmenhorst, E.-M., Hoermann, H.-J., Oeltze, K., Pennig, S., Rolny, V., Verjvoda, M., Staubach, M. & Schießl, C. (2013). Validierung eines Fitness-for-Duty Tests zur Steigerung der Sicherheit in Luftfahrt und Verkehr. Research Report DLR-Project "FIT". DLR: Cologne/Germany.

Hart, S.G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp 904–908). Santa Monica.

Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* . Amsterdam: North Holland Press.

Haslbeck, A., Eichinger, A., & Bengler, K. (2013). Pilot Decision Making: Modeling Choices in a Go-Around Situation. 17th International Symposium on Aviation Psychology, Dayton, Ohio.

Haslbeck, A., Kirchner, P., Schubert, E., & Bengler, K. (2014). A Flight Simulator Study to Evaluate Manual Flying Skills of Airline Pilots. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, pp 11–15). SAGE Journals.

Haslbeck, A., Schubert, E., Onnasch, L., Hüttig, G., Bubb, H., & Bengler, K. (2012). Manual flying skills under the influence of performance shaping factors. *Work: A Journal of Prevention, Assessment and Rehabilitation*, *41*(Supplement 1/2012), 178–183. Retrieved from http://dx.doi.org/10.3233/WOR-2012-0153-178.

Jacobson, N. S. & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59 (1),* 12–19.

Mueller, S. T. (2011). PEBL: The Psychology Experiment Building Language. Retrieved on Nov 8, 2011 from http://pebl.sourceforge.net/.

Pfendler, C. & Pitrella, F.D. & Wiegand, D. (1994). *Workload measurement in human engineering testing and evaluation*. Wachtberg: Forschungsinstitut für Anthropotechnik. Report Nr. 109.

Samel, A., Wegmann, H.M., & Vejvoda, M. (1995). Jet lag and sleepiness in aircrew. *Journal of Sleep Research, 4*, 30-36.

Samn, S.W., Perelli, L.P., 1982. *Estimating aircrew fatigue: a technique with application to airlift operations*. Brooks AFB, USAF School of Aerospace Medicine. Technical report SAM-TR-82-21.

Tritschler, K. & Bond, S. (2010). The influence of workload factors on flight crew fatigue. 63rd Annual IASS, *Information at a Glance*. Milan, Italy.

Van Dongen, H.P.A., Maislin, G., Mullington, J.M., Dinges, D.F. (2003). The cumulative cost of additional wakefulness: Dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, *26*, 117–126.

Williamson, A., Lombardi, D.A., Folkard, S., Stutts, J., Courtney, T.K., & Connor, J.L. (2011). The links between fatigue and safety. *Accident Analysis & Prevention*, *43*. 498-515.