

European Association for Aviation Psychology Conference EAAP 35

Results and implications of objective fatigue and performance measurements in five European Air Traffic Service Providers

Marsman, L.A.^a, Bos, T.^a, Verhoeven, R.^a, Cabon, P.^b, Vrancken, P.^b, & Drongelen van, A.^{a*}^aRoyal Netherlands Aerospace Centre (NLR), Anthony Fokkerweg 2, Amsterdam 1059CM, Netherlands^bWelbees, 207 Rue de Bercy, Paris 75012, France

Abstract

Air Traffic Controllers (ATCOs) face numerous fatigue-related challenges due to irregular working hours and demanding workload. The objective of this study was to determine how fatigue can be objectively monitored in the ops room and how they relate to subjective ratings. To do so, objective fatigue (eye tracking) and performance (PVT) data was gathered in five ATSPs (N=20). Alongside the objective measurements, fatigue (SP), sleepiness (KSS), and workload (ISA, RSME) questionnaires were administered. Measurements took place during ten so-called fatigue hotspot shifts, which were determined using both biomathematical and expert analyses. The results show that mean fatigue and sleepiness experienced during the shifts was moderate and below critical levels for a majority of the time. A clear increase in fatigue could be seen throughout shifts, as shown by the difference between mean subjective fatigue levels before and after duties, and corresponding PVT metrics. Significant correlations were found between subjective fatigue and sleepiness ratings on the one hand, and eye tracking metrics reflecting the percentage of time the eyes were closed for more than 70 or 80 percent (PERCLOS70 and 80) on the other hand. No significant correlations were found between subjective fatigue and sleepiness ratings and any of the PVT metrics. Furthermore, no significant correlations were found between the eye tracking metrics and the subjective workload ratings. The outcomes of this study underline that different types of fatigue-related outcome measures (subjective, objective, performance based) reflect different concepts and should be seen as complementary to one another when measuring fatigue in ATCOs. As such, relying on self-reported ratings alone may not offer a comprehensive understanding of individual fatigue levels and associated performance capabilities. The feasibility of using current objective measurement techniques by ATSPs themselves is however quite low, given the time, labour and operational constraints involved. Future research could thus target technological developments such as wireless EEG, fNIRS, speech analysis, and webcam-based eye tracking to assess fatigue in a more accessible and unobtrusive way.

* Corresponding author. Tel.: +31 88 511 36 27

E-mail address: Alwin.van.drongelen@nlr.nl

© 2024 The Authors. Published by ELSEVIER B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the European Association for Aviation Psychology Conference EAAP 35

Keywords: Air traffic control; fatigue; sleepiness; subjective measurements; objective measurements, validation

1. Introduction

Air Traffic Controllers (ATCOs) face numerous fatigue-related challenges due to irregular working hours and demanding workload. In response to the European regulations aimed at preventing ATCO fatigue and stress, a scientific evaluation funded by EASA was undertaken across member states. The objective of this study was to assess the level of fatigue risk and identify primary contributing factors within current ATSP operations. From the literature review carried out within the project it was shown that dedicated studies looking into the prevalence of fatigue among European air traffic controllers (ATCOs) are still lacking, especially those that attempt to objectively quantify the issue. The aim of the present study therefore was to validate subjective fatigue measurements of EU ATCOs during the ATC operation using objective measurement techniques. Next, the feasibility of using this equipment to measure fatigue in real-time in the ops room was determined.

2. Methods

2.1. ATSPs and participants

After an extensive selection procedure to find representative and eligible EU ATSPs, designated Points of Contact (PoCs) of five ATSPs were contacted, after which a meeting was set-up to explain the purpose and practicalities of the data collection: it was explained that at least four voluntarily participating ATCOs were needed to obtain valid information about the fatigue levels during shifts that were determined as fatigue hotspots, by means of both expert analyses and two biomathematical models. Next, the PoCs recruited volunteers and identified the dates on which they were working on the predefined shifts, and as such were selected to carry out the measurements.

Before the shifts, the participants were given more detailed information by means of a briefing presentation. Subsequently, they were asked to provide their informed consent. All participants were informed that their participation was completely voluntary and that they could revoke their consent to participate at any point. It was explained that all data is collected and processed in accordance with data protection legislation, and that the data is anonymized, digitalized, and saved on the secured networks. After providing their informed consent, participants were presented with a baseline questionnaire concerning demographics, experience, the quality of the workplace facilities, and fatigue and sleep characteristics. Subsequently, a briefing was held, either in the control room or a designated meeting room, discussing all procedures and materials related to the measurements during their shifts. Finally, the participants were instructed to practice the PVT to familiarize themselves with the device (tablet) and PVT application.

2.2. Procedure

For the study, a combination of objective and subjective measurements was used. The PVT was conducted both immediately before and after the designated shift. The eye-trackers were strategically positioned at two workstations, and recorded data throughout the entire shift. The subjective fatigue and workload ratings were collected before and after shifts, and approximately every hour during a shift using paper logs, with minor adjustments made based on shift length and breaks. Detailed logfiles were kept by the researchers each hour, during breaks, and at the shift completion.

2.3. Outcome measures

Eye tracking. The eye tracking (ET) outcomes blink duration, blink frequency, and the Percentage of eyelid Closure (PERCLOS; Abe et al., 2023) can be used as predictors for fatigue, which was found in both aviation (Peissl et al.,

2018) and the automotive industry (Zandi et al., 2019). Eye tracking has also been applied to measure workload, in which mainly blink frequency and blink duration are used as indicators (Benetto et al., 2011). PERCLOS, the proportion of time that the eyes were closed, was calculated for at least 70% or 80% of eye closure (PERCLOS70 and PERCLOS80; Abe, 2023), and was normalized per subject using the maximum eye-lid opening recorded. Because blinks tend to last longer when a person is fatigued, blinks were included in the PERCLOS calculations (Thropp et al., 2018; Abe et al., 2011). Blink duration is on average 100–400 milliseconds, while eye closure of >1000 ms has been defined as a microsleep (Bistow et al., 2005; Schiffman, 2001). Therefore, eye closure between 50 and 1000 ms was considered as a blink. Blink frequency was calculated as the average number of blinks per minute.

Remote eye trackers were used for the measurements as they tend to be less intrusive for the operators when used for longer durations. These eye trackers use optical tracking supported by infrared light. By means of corneal reflection, the eye trackers monitor pupil size, pupil shape and the position of the eyelids in real-time, which software computes into different outcomes measures. Two types of eye trackers from one manufacturer were used, the SmartEye Pro (consisting of separate cameras and flashers that can be flexibly positioned around an ATCO workstation) and the SmartEye Aurora (consisting of one bar with two integrated cameras and one flasher).

Psychomotor Vigilance Task (PVT). The PVT is a sustained-attention, simple reaction-time task that measures the speed at which subjects respond to a visual stimulus. The PVT has been extensively used in both laboratory and real-world settings, including aviation, to measure vigilance (e.g. Grant et al., 2017). While the PVT can be administered with several durations (3, 5 or 10 minutes), the 3-minute version was chosen for practical reasons. A dedicated PVT app on a standalone tablet was used to ensure that the measured reaction time (RT) was not influenced by the processing capacity of personal devices, thereby enhancing the validity of the outcome metric (Arthurs et al., 2021).

Subjective Measures. Subjective sleepiness and fatigue were measured using the Karolinska Sleepiness Scale (KSS) and the Samn-Perelli (SP) fatigue scale respectively. The KSS is a 9-point sleepiness scale (1 = extremely alert; 9 = extremely sleepy, great effort to keep alert, fighting sleep), which has been extensively validated (e.g. Åkerstedt & Gillberg, 1990). Sleepiness levels of 8 and 9 have been associated with high probability of accidents in simulators, for which it can be seen as a cut-off for critical fatigue. The SP is a 7-point fatigue scale (1 = fully alert, wide awake; 7 = completely exhausted, unable to function effectively) on which participants are asked for the level of fatigue they are currently experiencing. A cut-off of 6 and above has previously been used to indicate critical fatigue (Powell et al., 2011). Finally, two scales were selected to measure workload. The Rating Scale Mental Effort (RSME; Zijlstra, 1993) is a continuous scale ranging from 0 – 150 on which participants are asked to mark the point which they feel reflects their current workload level (0 = absolutely no effort; 110 = extreme effort). The Instantaneous Self-Assessment (ISA) scale provides a simple unambiguous workload measure with minimal intrusion and consists of a 5-point scale from 1 (under-utilised) to 5 (excessive) (Leggatt, 2005).

2.4. Data processing and analysis

After data collection, for the different types of data (ET, PVT, subjective questionnaires and baseline data) different data processing and analysis steps were applied. First, all information from the baseline questionnaire was entered into a separate datafile, using each participant's unique identification code. This data formed the basis for the subsequent analyses, while the identification code was used to match data per participant.

Next, the ET data was cleaned and prepared for analysis. Data intervals of 10 minutes were selected, corresponding to the moment the subjective ratings were filled out. Preferably, the 10-minute interval took place before the rating but in cases where no ET data was available before the rating (e.g. because a questionnaire was completed directly after a break), the 10-minute interval after this rating was selected. Next, within each slice of 10-minute ET data, the longest period, with a minimum of 1 minute, of good quality data was determined, extracted, and used to calculate the fatigue-related metrics (PERCLOS, blink duration and blink frequency). For each slice, the longest period of uninterrupted good quality data was selected. However, this period varied substantially, hence all slices with less than 60 seconds of data were excluded from analysis to ensure a more valid data set for subsequent analysis.

Second, all PVT data was transferred to NLR servers, after which the data of the different ATSPs was combined and categorised for ATSP, unit, shift type, and time during the shift (before or after). For each participant, a confidence interval (CI) was established to determine the normal range of RTs for each participant. All RTs outside of this CI were regarded as 'lapses'. Next, from the remaining data, the average RT and average speed (1/RT) was calculated

for each PVT test performed. These metrics were chosen to maximise sensitivity (see Basner & Dinges, 2011) due to substantial inter-participant variability in response times, as well as differences in logging speed between tablets.

Third, all subjective ratings were entered into a datafile. When needed, time zone corrections were made to match the clock times with those of the ET software. All demographic, function and shift-related information acquired was analysed using descriptive statistics. To determine if the chosen subjective measures were valid indicators for fatigue and workload in the ATC operation, all processed ET data was combined with the corresponding subjective ratings and uploaded in the statistical software IBM SPSS statistics (version 20) for further analysis. Subjective ratings for fatigue and sleepiness were correlated with blink duration, blink frequency, and PERCLOS70/80. Subjective ratings for workload were correlated with blink duration and blink frequency only, and Spearman correlations were performed on all scale and ordinal data. In order to find out if, and to what extent, the subjective ratings were associated with the chosen PVT outcome measures, this data was analysed using Spearman correlations as well.

Finally, a double multivariate repeated measures ANOVA was conducted to investigate the effect of time (before - after the shift) on the outcome measures. Any significant results were then verified using Wilcoxon Signed-Ranks tests. Finally, the effects of Shift (night vs other) and Unit (ACC, TWR) were analysed using Mann-Whitney U tests, with the deltas (after – before) of KSS, SP and mean PVT speed as dependent variables. For these outcome measures, a distinction was made between Area Control (ACC) and tower (TWR) units, as measurements were conducted in ACC and tower environments only. For shift type, several shift types (split, evening) were grouped into ‘other’ shift type, since a limited set of these shifts were measured.

3. Results

3.1. Participants

Demographics of the participating ATCOs can be found in Table 1. The participants worked at least one shift that was determined as a fatigue hotspot. In total, ten different shifts were measured, involving 20 participants. The average shift duration was approximately 8.5 hours (min=5 hours, max=11 hours).

Table 1. Demographics of 20 participating ATCOs (N=Number)

		n	Percentage (%)
Gender	Female	8	40.0
	Male	12	60.0
Age	20-29	3	15.0
	30-39	5	25.0
	40-49	10	50.0
	50-59	2	10.0
Unit Type	Area Controller (ACC)	12	60.0
	Tower Controller (ground or runway)	4	20.0
	Tower Controller (Approach)	4	20.0
Shift	Evening shift	6	30.0
	Night shift	12	60.0
	Split-shift (morning and afternoon)	2	10.0
Experience	<5 years	3	15.0
	6-10 years	2	10.0
	11-15 years	9	45.0
	16-20 years	3	15.0
	20+ years	3	15.0

3.2. Subjective ratings

In Table 2, the descriptive statistics of the subjective outcome measures ISA, RSME, KSS, and SP can be found. The average values for KSS and SP were below the cut-offs for high fatigue/sleepiness. The mean workload showed to be moderate to low (mean ISA 2.20, mean RSME 32.0). As for the comparison between pre- and post-duty values, the average KSS score increased from 3.96 before to 5.50 after the shift. For the SP, mean scores increased from 3.33 (before) to 4.31 (after). Significance testing is performed in Section 3.5.

Table 2. Descriptive statistics of the four subjective questionnaires used (N=Number, SD=Standard Deviation).

Outcome measure	N	Minimum	Maximum	Mean	SD
ISA (range 1-5)	105	1	4	2.20	0.74
RSME (range 0-150)	106	3	100	32.0	19.69
KSS (range 1-9)	118	1	8	4.56	1.63
SP (range 1-7)	118	2	6	3.75	0.89

3.3. Eye tracking data

In total, 123 data slices of eye tracking data could be derived from the 20 participants. After excluding the slices with less than 60 seconds, 106 slices of data remained. On average, 42.7% of the data from these data slices was usable as per quality requirements set (SD=26.2%, min=12%, max=99%), resulting in an average data duration of 259.2 seconds (SD=157.9, min=73, max=600) which equals 4 minutes and 29 seconds. Based on the extraction of the relevant ET data, the associated fatigue-related metrics were calculated for each slice (Table 3). All four ET measures show a large variation. PERCLOS70 for instance ranges between 0.47% and 45.25%. The average value for blink frequency is 12.95 blinks per minute (SD=8.28). For blink duration the average was 0.32 seconds (SD=0.05).

Table 3. Descriptives of the calculated eye tracking measures (N=Number, SD=Standard Deviation).

Outcome measure	N	Mean	Minimum	Maximum	SD
Blink duration (seconds)	106	0.32	0.22	0.55	0.05
Blink frequency (blinks per minute)	106	12.95	0.51	40.49	8.28
PERCLOS70 (percentage of time)	106	7.93	0.47	45.25	9.23
PERCLOS80 (percentage of time)	106	3.84	0.19	26.11	4.11

3.4. PVT

Table 4 presents the descriptive statistics for the PVT outcome measures both before and after duty. No data were excluded from the analysis. Four participants were measured during two consecutive duties, for which the total number of recordings was 24. As can be seen, there was an increase in all outcome measures between the measurements that were performed before (pre-duty) and after the fatigue hotspot shift (post-duty). This is also shown for speed and the number of lapses. **Error! Reference source not found..** Significance is tested in Section 3.5.

Table 4. Descriptive statistics of the Psychomotor Vigilance Task (PVT) before and after duty (N=Number, SD=Standard Deviation).

Outcome measure	Pre-duty					Post-duty			
	N	Min	Max	Mean	SD	Min	Max	Mean	SD
RT (ms)	24	290.48	700.19	389.65	86.11	311.23	714.65	403.27	97.86
Speed (m/s)	24	1.43	3.44	2.67	0.47	1.40	3.21	2.59	0.48
Lapses (N)	24	0.62	1.46	1.04	0.20	0.71	2.377	1.54	0.40

3.5. Effect of time (pre- and post-duty) and duty type

An repeated-measures ANOVA showed that Time had a significant effect on at least one of the dependent variables (multivariate Wilks' $F(4,20)=4.633$, $p=.008$). The univariate contrasts tests showed that this particularly concerned a significant and a strong change in the KSS ($F(1,23)=13.125$, $p=.001$, $\eta^2=.363$) and the SP ($F(1,23)=15.018$, $p=.001$, $\eta^2=.395$) between the after versus before duty measurements (Figure 2). The significant result of Time on KSS ($Z=-3.193$, $p=.001$) and SP ($Z=-3.107$, $p=.002$) values was confirmed with two Wilcoxon Signed-Ranks tests. The effect of Time on the PVT outcome measures was non-significant. Next, two Mann-Whitney U tests were conducted to test the effects of Shift type and Unit. This test showed no significant differences between the night shift and other shifts regarding the outcome measures. Shift type thus did not have an effect on both the subjective and objective outcome measures. However, the Mann-Whitney U test with Unit (ACC, TWR) as the grouping variable did show a significant difference between ACC and TWR with respect to their mean rank of the delta of speed ($U=31.00$, $p=.043$). This means that the PVT outcome measure speed after the duty worsened significantly more in ACC compared to TWR.

3.6. Relationship between eye tracking metrics and subjective ratings

The Spearman's rank correlations were used to test the relationship between subjective measures (workload and fatigue) and the ET and PVT measures. A significant correlation was found between workload (ISA and RSME) scores ($r=0.653$, $p<.001$) which indicates the workload rating scales used both reflect the experienced workload by the ATCOs. Next, a high correlation was found between KSS and SP ($r=0.796$, $p<.001$), showing that subjective fatigue and subjective sleepiness is closely related. For the ET metrics, blink duration and blink frequency did not prove to be correlated with each other, but PERCLOS70 and PERCLOS80 were ($r=0.944$, $p<.001$). Regarding the correlation between the subjective workload ratings and the objective ET metrics no significant results were found. The blink metrics did not correlate with sleepiness and fatigue ratings either. A moderate significant correlation was found between both SP and KSS ratings and percentage eye closure (PERCLOS70) ($r=.237$, $p=.017$ and $r=.204$, $p=.041$ respectively). For sleepiness, a moderate significant correlation with PERCLOS80 was found as well ($r=.203$, $p=.042$). No significant correlations were found between any of the subjective ratings and any of the PVT metrics.

4. Discussion

The aim of this study was to validate subjective fatigue and workload outcomes by means of objective measures (being eye tracking and PVT). Data was collected in five European ATSPs, using 20 voluntarily participating ATCOs during ten different fatigue hotspot shifts. The participants consisted of both males and females of different types of units, age categories and levels of experience. Measurements were conducted during night shifts mostly (60%) as they were identified as fatigue hotspots most often, but during evening (30%) and split shifts (10%) as well. The subjective measurements during these shifts showed that the mean fatigue experienced was moderate and below critical values for the majority of the time. Mean workload showed to be moderate to low, which could have been a result of the night duties: in some ATSPs there was still substantial traffic during the night, while in others only occasional traffic needed to be handled. Nevertheless, a significant increase in fatigue could be seen throughout the shift. ET data proved to be of reasonable quality, although of the ten minutes intervals that were selected, only a mean duration of 4.5 minutes of continuous data could be used for the analysis. The main reason for ET data to be excluded was a lack of detection of the eyes by the eye tracker during those intervals, for instance because the ATCO looked away from the navigation screen. All four eye tracking metrics that were derived from the data showed a large variation, most probably because of both between and within participant differences.

No significant correlations were found between the subjective fatigue and sleepiness ratings and the PVT metrics. Although the correlations found were pointing in the right direction (the higher the fatigue/sleepiness ratings, the higher the RT, the lower the speed, and the higher the number of lapses), none of these reached statistical significance. Although this may be caused by the small sample, subjective ratings of sleepiness and fatigue do not necessarily correspond with sustained alertness as measured by a reaction time task either, hence other factors, such as individual differences in cognitive processes, may influence this relationship. Previous studies have reported mixed results for the correlation between PVT metrics and subjective fatigue (Van Dongen et al., 2011; Wehrens et al., 2012) and this ambiguous association was also found in ATCO specific literature. While Li et al. (2021) did find an association of fatigue with performance, Chen et al. (2019) found a discrepancy between sleepiness ratings and objective PVT data. Reaction time tasks and subjective fatigue/sleepiness measures therefore seem to capture different concepts and, if possible, should be used alongside each other in (ATC) research concerning fatigue and performance.

In addition, the results of this study indicated that blink frequency and duration do not seem to be adequate measures to detect ATCO workload through eye tracking in the ops room. No significant correlations were found between these ET metrics and the subjective workload ratings (RSME, ISA). Although blink frequency and blink duration have been shown to be related to workload (Benetto et al., 2011), this concerned driving in a simulator and it could be possible that these results are therefore not transferable to ATC. In addition, the direction of the correlation between blinks and workload is not always clear: while multiple studies have found a decrease in blink frequency as a result of increased workload (Veltman et al., 1998; Zheng et al., 2012), others found an increase in blink frequency as task load increased (Magliacano et al., 2020; Tsai et al., 2007). In addition, when operators suffer from drowsiness, a significant increase of blink frequency has been found (Hu & Lodewijks, 2020). As a result of its multidimensional nature, using blink metrics as indicator for both fatigue and workload could be problematic. Since fatigue can be related to sleep/extended

wakefulness, underload, and overload (causing mental fatigue and distress), this could also explain the non-significant correlations between the subjective fatigue ratings and both blink frequency and duration found. Other ET parameters, derived from pupillography (e.g. pupil diameter) may be better able to capture experienced workload, although illumination should then be kept as constant as possible (Marquardt et al., 2015).

Significant correlations were found between subjective fatigue and sleepiness ratings and eye tracking metrics (PERCLOS70 and 80). Although the correlations found were moderate (r between 0.20 and 0.25), this finding substantiates the outcomes of previous studies, indicating that the time during which the eyes are closed for a certain percentage (e.g. 70%) increases with the onset of fatigue (e.g. Zhang et al., 2021). Hu and Lodewijks (2020) stated that PERCLOS is the most reliable and valid indicator of fatigue since this metric was shown to increase with the extension of wakefulness, and is also associated with performance decrements. The subjective measures used (predominantly KSS) therefore seem to be good indicators for ATCO fatigue and sleepiness in the ops room.

4.1. Strengths, limitations, and implications for future research and practice

For the ET measures, the choice was made to calculate PERCLOS70 and 80, as these are the most common. However, it could be argued that other measures (e.g. PERCLOS 60) could have been used as well, since with a lower eye closure percentage as an additional outcome metric, a different (stronger) correlation might have been found (Zhang et al., 2021). In addition, for the current analyses, all slices with ET data that had less than 60 seconds of continuous data of good quality were excluded from analyses to ensure the inclusion of valid measurements only, and minimise noise in the dataset. However, in previous studies, longer intervals of data have been used. Sommer and Golz (2010) for example took slices of three-minutes from a larger ET data set, and found a strongly decreased correlation coefficient with subjective outcome measures, in comparison with using all data. Furthermore, in the current study it was chosen to include blinks in the PERCLOS calculations, whereas in other studies, eye closures that last less than one second were excluded (e.g. Abe et al, 2020).

Although there seem to be no golden values regarding the cut-offs to be used in ET analyses, one needs to be aware that the choices made while preparing the data can substantially influence the results. As a result, more research is needed to discover best practices for the usage of eye tracking in the (ATSP) workplace. Regardless of methods, the fatigue and sleepiness levels obtained during the measurements may have influenced the outcomes of the present study. The fatigue hotspot shifts were selected to detect relatively large changes in ATCOs' fatigue. However, subjective fatigue was found to be moderate throughout the largest part of shifts. From an operational point of view this is reassuring, but from a scientific perspective the lack of variation and moderate levels of fatigue could explain the relative low correlations found between subjective and objective outcome measures.

The measurements conducted proved to be quite time consuming and laborious. Because of the organisation and expertise needed, and the burden for ATCOs, it will therefore be difficult to implement eye tracking in its current form in daily ATSP operations. Nevertheless, this study did show that studying fatigue this way (e.g. for scientific purposes and/or by external experts) is feasible during a limited period, and can result in good quality ET and PVT data in the ops room. Given these limitations, one should consider advancements in other non-intrusive (psycho)physiological measurement techniques (e.g. functional near-infrared spectroscopy (fNIRS) or speech) for ATCO fatigue detection in the future. Furthermore, future eye tracking through standard video footage (e.g. a webcam) would offer a non-invasive and cost-effective alternative compared to the relatively complex eye tracking systems used nowadays. The implementation of these techniques raise concerns related to ethics and data privacy however, particularly in the context of GDPR guidelines. Future developments in fatigue detection should address these (ethical) considerations to gain wider acceptance within the ATC community.

4.2. Conclusions

The objective of this study was to determine how fatigue could be objectively monitored in the ops room and how subjective ratings are associated with these measures. Significant correlations were found between subjective fatigue and sleepiness ratings on the one hand, and ET metrics reflecting the percentage of time the eyes were closed on the other hand. No significant correlations were found between the subjective fatigue and sleepiness ratings and any of the PVT metrics. Furthermore, no significant correlations were found between the ET metrics and the subjective

workload ratings. The outcomes of this study underline that the different types of fatigue-related outcome measures (subjective, objective, performance based) reflect different concepts and should be seen as complementary to one another when measuring fatigue in ATCOs. As such, relying on self-reported ratings alone may not offer a comprehensive understanding of the individual fatigue levels and associated performance capabilities.

References

- Abe, T. (2023). PERCLOS-based technologies for detecting drowsiness: current evidence and future directions. *Sleep Advances*, 4(1), zpad006.
- Abe, T., Nonomura, T., Komada, Y., Asaoka, S., Sasai, T., Ueno, A., & Inoue, Y. (2011). Detecting deteriorated vigilance using percentage of eyelid closure time during behavioral maintenance of wakefulness tests. *International Journal of Psychophysiology*, 82(3), 269-274.
- Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52(1-2), 29-37.
- Arthurs, M., Dominguez Veiga, J. J., & Ward, T. E. (2021, September). Accurate reaction times on smartphones: The challenges of developing a mobile psychomotor vigilance Task. In *Proceedings of the 2021 ACM International Symposium on Wearable Computers* (pp. 53-57).
- Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep*, 34(5), 581-591. <https://doi.org/10.1093/sleep/34.5.581>
- Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., & Montanari, R. (2011). Driver workload and eye blink duration. *Transportation research part F: traffic psychology and behaviour*, 14(3), 199-208.
- Bristow, D., Haynes, J. D., Sylvester, R., Frith, C. D., & Rees, G. (2005). Blinking suppresses the neural response to unchanging retinal stimulation. *Current biology*, 15(14), 1296-1300.
- Chen, M. L., Lu, S. Y., & Mao, I. F. (2019). Subjective symptoms and physiological measures of fatigue in air traffic controllers. *International Journal of Industrial Ergonomics*, 70, 1-8.
- Grant, D. A., Honn, K. A., Layton, M. E., Riedy, S. M., & Van Dongen, H. P. (2017). 3-minute smartphone-based and tablet-based psychomotor vigilance tests for the assessment of reduced alertness due to sleep deprivation. *Behavior research methods*, 49, 1020-1029.
- Hu, X., & Lodewijks, G. (2020). Detecting fatigue in car drivers and aircraft pilots by using non-invasive measures: The value of differentiation of sleepiness and mental fatigue. *Journal of Safety Research*, 72, 173-187.
- Leggatt, A. (2023). Validation of the ISA (Instantaneous Self Assessment) subjective workload tool. In *Contemporary Ergonomics 2005* (74-78). Taylor & Francis.
- Li, W.-C., Kearney, P., Zhang, J., Hsu, Y.-L., & Braithwaite, G. (2021). The analysis of occurrences associated with air traffic volume and air traffic controllers' alertness for fatigue risk management. *Risk Analysis*, 41(6), 1004-1018.
- Magliacano, A., Fiorenza, S., Estraneo, A., & Trojano, L. (2020). Eye blink rate increases as a function of cognitive load during an auditory oddball paradigm. *Neuroscience Letters*, 736, 135293.
- Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of eye-related measures of drivers' mental workload. *Procedia Manufacturing*, 3, 2854-2861.
- Powell, D., Spencer, M. B., & Petrie, K. J. (2011). Automated collection of fatigue ratings at the top of descent: A practical commercial airline tool. *Aviation, Space, and Environmental Medicine*, 82(11), 1037-1041.
- Samn S, & Perelli L. (1982). Estimating aircrew fatigue: A technique with implications to airlift operations. Technical Report No. SAM-TR-82-21. Brooks A.F.B.TX: USAF School of Aerospace Medicine, 19821-26.
- Schiffman, H.R., (2001) *Sensation and Perception. An Integrated Approach*, NY: John Wiley and Sons, Inc.
- Smith, P. A., Di Milia, L., Smith, A., Gee, W., & Mackay, R. (2000). Managing fatigue through roster design. In *Proceedings of The Queensland Mining Industry Health and Safety Conference: A New Era in Mine Health and Safety Management*.
- Sommer, D., & Golz, M. (2010). Evaluation of PERCLOS based current fatigue monitoring technologies. *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*.
- Thropp, J.E., Scallon, Jeffrey F. V., Buza, P. (2018), PERCLOS as an indicator of slow-onset hypoxia in aviation, *Aerospace Medicine Human Performance*. 2018; 89 (8):700–707.
- Tsai, Y. F., Viirre, E., Strychacz, C., Chase, B., & Jung, T. P. (2007). Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviation, Space, and Environmental Medicine*, 78(5), B176-B185.
- Wehrens, S. M. T., Hampton, S. M., Kerkhofs, M., & Skene, D. J. (2012). Mood, Alertness, and Performance in Response to Sleep Deprivation and Recovery Sleep in Experienced Shiftworkers Versus Non-Shiftworkers. *Chronobiology International*, 29, 537–548.
- Van Dongen, H. P. A., Caldwell, J. A., & Caldwell, J. L. (2011). Chapter 9—Individual differences in cognitive vulnerability to fatigue in the laboratory and in the workplace. In H. P. A. Van Dongen & G. A. Kerkhof (Eds.), *Progress in Brain Research* (Vol. 190, 145–153).
- Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of Task difficulty. *Ergonomics*, 41, 656–669.
- Zandi, A. S., Quddus, A., Prest, L., & Comeau, F. J. (2019). Non-intrusive detection of drowsy driving based on eye tracking data. *Transportation Research Record*, 2673(6), 247-257.
- Zhang, J., Chen, Z., Liu, W., Ding, P., & Wu, Q. (2021). A Field Study of Work Type Influence on Air Traffic Controllers' Fatigue Based on Data-Driven PERCLOS Detection. *International Journal of Environmental Research and Public Health*, 18(22), Article 22.
- Zheng, B., Jiang, X., Tien, G., Meneghetti, A., Panton, O. N. M., & Atkins, M. S. (2012). Workload assessment of surgeons: Correlation between NASA TLX and blinks. *Surgical Endoscopy*, 26(10), 2746–2750.
- Zijlstra, F.R.H. (1993). Efficiency in work behavior. A design approach for modern tools. PhD thesis, Delft University of Technology, Delft University Press, Delft, The Netherlands.