Article

Air Traffic Controllers' Rostering: Sleep Quality, Vigilance, Mental Workload, and Boredom: A Report of Two Case Studies

Michela Terenzi , Giorgia Tempestini , Francesco Di Nocera *

Posted Date: 17 May 2024

doi: 10.20944/preprints202405.1120.v1

Keywords: fatigue; vigilance; workload; boredom; rostering

Article

Air Traffic Controllers' Rostering: Sleep Quality, Vigilance, Mental Workload, and Boredom: A Report of Two Case Studies

Michela Terenzi [1], Giorgia Tempestini [2] and Francesco Di Nocera [2,*]

[1] Deep Blue Srl, 00185 Rome, Italy
[2] Department of Planning, Design, and Technology of Architecture, Sapienza University of Rome, 00196 Rome, Italy
* Correspondence: francesco.dinocera@uniroma1.it

Abstract: Fatigue is a recognized concern in air traffic management, prompting international bodies like ICAO and EASA to advocate for Fatigue Risk Management Systems (FRMS). EU Regulation 2017/373, effective since January 2020, imposes specific requirements on air traffic service providers regarding controllers' fatigue, stress, and rostering systems, as part of safety management protocols. Despite active campaigns promoting fatigue awareness in air traffic control (ATC), there remains a need for standardized operational requirements. Eurocontrol recently published "Guidelines on fatigue management in ATC rostering systems" (April 23), aiming to facilitate the adoption of common fatigue operational standards. However, neither EU Regulation 2017/373 nor existing documentation specifies exact rostering criteria. Air Navigation Service Providers (ANSPs) typically derive these criteria from scientific literature, best practices, historical data, and legal and operational necessities. Yet, assessing and monitoring fatigue in operational settings can be complex and require long-term studies due to its multifaceted nature, spanning factors such as sleep quality, circadian rhythms, psychosocial influences, individual traits, and environmental conditions. The case studies reported in this paper represent an attempt to create an evidence-based protocol for fatigue risk monitoring in ATC operations, utilizing a non-invasive approach and collecting multidimensional data. The two case studies involved en-route and tower controllers from different ATC Centers, and the results underscore the importance of fatigue assessment in ATC and elucidates challenges in implementing fatigue monitoring within operational contexts.

Keywords: fatigue; vigilance; workload; boredom; air traffic controller; rostering

---

1. Introduction

Air traffic controllers are required to manage particularly complex and dynamic situations quickly, aiming to identify and resolve potential conflicts and optimize flight paths without compromising safety. Therefore, they must perceive, understand, and anticipate many aircraft characteristics and flight paths. These types of tasks require the involvement of various resources (both mental and physical) by individuals. They may be subject to deterioration such that the performance of operators and, consequently, the safety of air traffic control and management operations is adversely affected. In addition to the time-consuming resources to be invested in the work activity, another critical issue is shift work: most air traffic controllers (ATCOs) in strategic ATC Centers have to cover uninterrupted shifts for 24 hours, 7 days a week.

In these operational contexts, workload, situation awareness, and vigilance are recognized as elements that can negatively influence operators' performance levels, affecting response times and the number of errors [1]. To control the impact of these phenomena on the entire system, a regulatory framework establishes measures to detect, monitor, and manage fatigue. European Commission Regulation (ER) 2017/373 ATS.OR.135 contains requirements and indications for developing a fatigue management policy, a procedure for identifying, preventing, mitigating, and monitoring fatigue levels. Also, it includes references to appropriate training and support for ATCOs personnel [2].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 May 2024 doi:10.20944/preprints202405.1120.v1

In addition to regulatory aspects aimed at controlling the level of fatigue in those performing safety-related activities, fatigue has been the subject of analysis in many research areas. Terenzi, Ricciardi, and Di Nocera [3] recently reviewed the literature intending to provide guidance for proper shift planning of ATCOs, primarily focusing on choosing appropriate parameters for defining workers' rest periods. Three elements were considered: 1) task-specific characteristics, 2) physiological needs of the operator, and 3) definition of rest periods within the shift schedule. Indeed. several studies have shown that, from a physiological point of view, night shifts result in reduced vigilance and sleep deprivation [4]. In terms of task characteristics, on the other hand, the high workload and monotony of some tasks provide little opportunity for the operator to implement compensatory strategies, making them more vulnerable to the effects of fatigue [5] In contrast, tasks with high levels of challenge and intrinsic interest can resist fatigue [6] The review provided guidance regarding the maximum shift duration (7-10 hours) to be evaluated according to the rotations and the rest following the night shift, which should be longer, considering that night hours have more significant implications on the operator's perception of fatigue.

Despite the numerous studies identifying requirements for shift scheduling in air traffic controllers, how fatigue is assessed and monitored still needs to be standardized. For this reason, the main objective of the studies reported here was to investigate the consequences of fatigue, seeking to understand, through an analysis of sleep, vigilance, mental workload, and boredom, what challenges are associated with fatigue monitoring within an operational context such as that of air traffic control. Results will contribute to the refinement of fatigue management strategies and facilitate future research and practical applications.

2. Rationale and Instruments

Two case studies were carried out to evaluate the professional fatigue and the level of alertness of air traffic controllers in the context of the current rostering system. The studies had two main objectives: 1) first, to assess the functional state of ATCOs through a set of subjective and objective indicators and then to provide an insight on the effectiveness of the rostering system in place; 2) secondly, to provide critical elements for the development of a monitoring protocol.

Case study 1 involved controllers working respectively in one Area Control Centre (ACC) (Group 1), where they provide assistance to aircraft in the en route phase of flight, whether they are landing at an airport or simply flying through the airspace while the other is a Control Tower (Group 2) of one high-density airport from which controllers handle take-offs, landings, and ground movements. Case study 2 involved operators working at two Control towers, one high-density (Group 3), one low-density (Group 4), and another ACC (Group 5).

We assessed all the operators during different work shifts (i.e., morning, afternoon, and night). Particularly, on each shift, we requested the operators to provide information about the duration and quality of their sleep, to provide a self-assessment of their mental workload and their boredom, and to perform a vigilance task before and after the work session. The instruments used in the following studies were the following:

- *Sleep log.* The sleep log was a set of questions to be filled out daily upon awakening. It allowed the recording of information regarding sleep duration and quality for each sleep period.
- *Psychomotor Vigilance Task (PVT)* [7]. The PVT is a few-minute, reaction time-based test designed to assess the ability to maintain attention and respond promptly to certain stimuli or cues. In the PVT, participants must react (by pressing a button) as quickly as possible to a visual stimulus on a device screen. The reaction time between the appearance of the stimulus and the participant's response is recorded and used to measure their alertness and speed of information processing. The software was installed on a smartphone and left at the working site available to the operators for the duration of the data collection period.
- *NASA Task Load Index (NASA-TLX)* [8]. The NASA-TLX is the most widely used subjective tool for estimating mental workload in various operational and experimental situations. The NASA-TLX requires the operator to rate along six scales (Mental Demand, Temporal Demand, Performance, Effort, Frustration) from 0 to 100 that collectively provide a total index of mental workload related to a specific task.
- *The Multidimensional State Boredom Scale* [9]. This instrument consists of 29 items to estimate the operator's boredom experience at completion. The MSBS consists of five subscales: Disengagement,

High Arousal, Low Arousal, Inattention, and Time Perception. The total score if commonly used to estimate boredom.

Sleep monitoring was carried out daily. Perceived workload (NASA-TLX) and boredom (MSBS) were monitored at the end of the work session. Vigilance (PVT) was observed before and after the work session.

3. Case Study 1

Two ATC Centers were involved in the study, hereinafter Group 1 and Group 2. The ACC handling the en-route traffic had a rostering schedule defined by three eight-hour shifts (morning, afternoon, and night). The Control Tower responsible for the traffic management of the high-density airport (more than 12.000 movements/y in 2021, when the study was conducted) had the same rostering schedule. Both ATC Centers operate 24 hours a day, 7 days a week (24/7).

3.1. Participants

Eighteen controllers belonging to Group 1 and Group 2 (see above) volunteered in this study. After data inspection, 5 participants with too much missing data in the questionnaires or who had not worked at least one shift in each time slot (morning, afternoon, night) were excluded from the sample. Thus, the final group consisted of 13 controllers (9 ATCOs and 4 TWR; mean age=48, sd=5.56; 7 females and 6 males). Operators were contacted based on the planning to cover all the shifts at least three times.

3.2. Procedure

At an initial meeting, instructions were provided, and informed consent was gathered from the controllers. The study's objective was explained to the ATCOs during this first meeting. They were given the materials, familiarized with the PVT task, and received an explanation of the questionnaires to be filled. Participants were informed that the study would last for 20 days, during which they had to fill out the sleep log journal daily, while they had to take the other tests only at work. To limit the invasiveness, we instructed the controllers and asked them to handle the self-administration of the tests.

3.3. Data Analysis and Results

Operators were variably diligent with respect to filling out the sleep log and end-of-shift questionnaires. In some cases, PVT data were missing either at the beginning or end of the shift. The final dataset contained several missing data. However, since multiple measurements were taken over time for the same shift, the mean values were computed over multiple points in almost all cases. ATCOs and TWR data were analyzed together to increase the sample size. However, all the following tables report the breakdown means and standard deviations for the two subgroups.

Sleep log journal. Sleep duration and quality were analyzed through repeated-measures ANOVA designs employing Shift (Morning vs. Afternoon vs. Night) as a factor. Results showed a statistically significant difference in sleep duration ($F_{2,10}$=32.92, p<.001). Specifically, post-hoc tests showed that when ATCOs must serve at night, they report sleeping less than when they have to serve in the morning and afternoon. The sleep quality shows an identical pattern ($F_{2,10}$ =4.81, p<05).

Table 1. Average sleep duration by Shift and Group in minutes (standard deviations in brackets).

	MORNING	AFTERNOON	NIGHT
GROUP 1	359 (52)	473 (78)	228 (82)
GROUP 2	380 (68)	421 (29)	388 (234)

Table 2. Average sleep quality by Shift and Group (standard deviations in brackets).

	MORNING	AFTERNOON	NIGHT
GROUP 1	3.47 (0.72)	3.92 (0.95)	2.49 (0.84)
GROUP 2	3.52 (1.12)	3.50 (0.71)	3.40 (1.40)

PVT. The difference between reaction times generated in test performance pre and post-working sessions was used in a repeated measures ANOVA design using Shift (Morning vs. Afternoon vs. Night) as a factor. No statistically significant difference was found ($F_{2,10}$ =.33, p>.05).

Table 3. Median PVT reaction time differences (pre-post working session) by Shift and Group (ranges in brackets).

	MORNING	AFTERNOON	NIGHT
GROUP 1	3.75 (-29 - 12)	3.50 (-19 - 43)	0 (-140 - 24)
GROUP 2	1.57 (-13 - 31)	0 (-11 - 6)	-55 (-87 - 5)

NASA-TLX. The total score at NASA-TLX was analyzed by a repeated measures ANOVA design employing Shift (Morning vs. Afternoon vs. Night) as a factor. No statistically significant difference was found ($F_{2,10}$ =.09, p>.05).

Table 4. Average NASA-TLX score by Shift and Site (standard deviations in brackets).

	MORNING	AFTERNOON	NIGHT
GROUP 1	34.22 (13.59)	24.24 (9.95)	26.91 (13.22)
GROUP 2	30.57 (16.22)	18.75 (3.26)	27.38 (12.25)

MSBS. MSBS scores were analyzed by repeated measures ANOVA design employing Shift (Morning vs. Afternoon vs. Night) as a factor. No statistically significant difference was found (F(2,18)=,02390, p>.05).

Table 5. Perceived boredom by Shift.

	MORNING	AFTERNOON	NIGHT
GROUP 1	1.78 (0.99)	1.97 (0.85)	1.84 (0.75)
GROUP 2	2.16 (0.75)	2.11 (0.72)	2.37 (0.79)

4. Discussion Case Study 1

Case Study 1 showed substantial homogeneity in workload, boredom, and vigilance regardless of shift (morning, afternoon, or night). Vigilance was also unaffected by the work session. That excellent result indicates that the implemented rostering does not affect any of the variables considered. Instead, the night shift appears to affect the duration and quality of sleep. However, it should be specified that ATCOs (especially TWRs) often worked double shifts (morning and night). Therefore, the values reported in the sleep log for the night before the morning shift sometimes were the same for the night shift. Although we attempted to stick as far as possible to the experimental rigor, this is a case study with all its limitations. The dataset is affected by missing data, and the results of this first case study principally served as lessons learned to foster a more structured approach, eventually leading to a systematic approach to fatigue monitoring.

The more important lesson learned from Case Study 1 was that data collection based on a comparable temporal distribution of the shifts was indeed an obstacle. That top-down organization led to "forcing" the ATCOs to dedicate time for data collection on specific dates. Perhaps it would be more beneficial to allow ATCOs to choose the days on which to run tests and stop the data collection when the necessary conditions are met (e.g., data collected on three days of morning, afternoon, and night shifts for a total of nine days, not necessarily consecutive). In this way, the ATCOs would choose the pacing of data collection and more reliable measurements could be obtained, albeit not collected consecutively. That is precisely what was implemented in Case Study 2 while keeping the rationale and the instruments used in Case Study 1 equal.

5. Case Study 2

One high-density airport (more than 75.000 movements/y in 2022) (Group 3), one low-density airport (in the 2022 ~ 1.000 movements/y) (Group 4) and one ACC (Group 5) were involved in the assessment of the Case Study 2.

The Control Tower of the high-density airport had a three, eight-hour shift roster (morning, afternoon and night) and 24/7 operating hours. While, the Control Tower of the low-density airport (Group 4) had a two, eight-hour shift rostering schedule (morning 07:00- 15:00, afternoon 15:00-23:00). No night shift was foreseen in this Centre, indeed the operating hours is 16/7. Finally, the ACC (Group 5) handling the en-route traffic had a rostering schedule defined by the usual three eight-hour shifts (morning, afternoon and night) and the operative hours is 24/7.

5.1. Participants

Twenty-four ATCOs from the three Air Traffic Control entities reported above (Group 3 = 10, Group 4 = 4, and Group 5 = 10; see section 2) participated in this study.

5.2. Procedures: Changes from the First Study

The current work integrates the lessons learned from Case Study 1 regarding the procedure, including the presentation of the study, data collection, and experimental plan. The study was presented through individual in-person meetings to share the purpose and methodology of the study. When in-person meetings were not feasible, remote meetings were organized with the same objectives.

For the data collection and experimental plan, an ecological approach was adopted because of the complexity of the current rostering system. ATCOs were individually responsible for taking the tests, following the protocol, and choosing suitable days within the set period. ATCOs to choose the days on which take the tests and the data collection stopped after three occurrences of morning, afternoon, and night shifts (nine days total). This approach allowed ATCOs to adjust the pace of data collection, resulting in fewer missing data.

Periodic meetings were held with the three agencies in coordination with ANACA. These meetings involved operational staff designated as the study's Point of Contact (POC) to monitor the progress of data collection.

5.3. Data Analysis

As for Study 1, the collected data were analyzed using ANOVA repeated measures designs by Shift (Morning vs. Afternoon vs. Night), employing all measures as dependent variables. Analyses were conducted on all ATCOs to increase sample size. However, all the following tables report the breakdown means and standard deviations for the two subgroups. Given its specific rostering schedule (see section 2) Group 4 did not contribute to measures related to night shifts.

Sleep log journal. The repeated measure ANOVAs conducted employing sleep duration and quality reported by ATCOs in the sleep log journal both showed a statistically significant effect [$F_{2,22}$=32.877, p<.001 and $F_{2,22}$=17.598, p<.001, respectively]. Duncan post-hoc testing showed that these effects were substantially due to the difference between night shift and the other two shifts.

Table 6. Average sleep duration by Shift and Group in minutes (standard deviations in brackets).

	MORNING	AFTERNOON	NIGHT
GROUP 3	357 (50)	447 (30)	240 (163)
GROUP 4	347 (22)	469 (46)	N/A
GROUP 5	378 (48)	414 (55)	214 (92)

Table 7. Average sleep quality by Shift and Group (standard deviations in brackets).

	MORNING	AFTERNOON	NIGHT
GROUP 3	3.09 (0.48)	3.63 (0.46)	2.30 (1.30)

GROUP 4	3.50 (0.64)	4.2 (0.58)	N/A
GROUP 5	3.21 (0.38)	3.39 (0.48)	2.43 (0.98)

Multidimensional State Boredom Scale. The repeated measures ANOVA conducted on the total scores on the Multidimensional State Boredom Scale showed no significant differences as a function of the shift [$F_{2,32}$=1.6551, p>.05].

Table 8. Average MSBS score by Shift and Site (standard deviations in brackets).

	MORNING	AFTERNOON	NIGHT
GROUP 3	1.98 (0.99)	2.12 (1.22)	1.95 (1.03)
GROUP 4	1.53 (0.49)	1.53 (0.53)	N/A
GROUP 5	2.53 (0.99)	2.66 (1.21)	2.74 (1.43)

NASA - Task Load Index. The repeated ANOVA conducted on total scores to the NASA - Task Load Index showed no significant differences as a function of shift [$F_{2,32}$=.31734, p>.05].

Table 9. Average NASA-TLX score by Shift and Site. Standard deviations in brackets.

	MORNING	AFTERNOON	NIGHT
GROUP 3	56.07 (16.59)	56.37 (16.61)	50.79 (16.92)
GROUP 4	26.45 (7.80)	29.64 (7.88)	N/A
GROUP 5	58.22 (14.64)	58.54 (13.87)	59.49 (20.71)

Psychomotor Vigilance Task. The repeated measure ANOVA conducted on the difference in reaction times to the Psychomotor Vigilance Task performed before and after the work session showed that there were significant differences by shift [$F_{2,32}$=3.7291, p<.05]. Duncan's post-hoc test showed that the difference was due to the night shift producing higher delta.

Table 10. Median PVT reaction time differences (pre-post working session) by Shift and Site (ranges in brackets).

	MORNING	AFTERNOON	NIGHT
GROUP 3	-9.44 (-133 - 32)	4.57 (-196 - 39)	-25.50 (-336 - 67)
GROUP 4	3.53 (-3 - 19)	6.79 (-3 - 16)	N/A
GROUP 5	-4,45 (-29 - 22)	0.99 (-177 - 35)	-25.18 (-223 - 36)

6. Discussion Case Study 2

The inherent nature of field studies, particularly those relying on self-reported data, often involves trade-offs between completeness, rigor, and ecological validity. While the ATCOs largely complied with data collection requests, some events (e.g., illness, leaves for training) had an impact. Despite the adjustments made according to the lesson learned from Case Study 1, also in this case questionnaires were not always filled out, and tests were not always performed. That resulted in missing data in the final dataset.

Case Study 2 confirmed substantial homogeneity in workload and boredom regardless of shift (morning, afternoon, or night). Vigilance was unaffected by the work session only in the morning and afternoon shifts. Results showed that a slight decrement in vigilance occurs on the night shift. As for Case Study 1, the results showed that the duration and quality of sleep reported by ATCOs varied according to shift. The effect is basically due to the difference between the night shift and the other two. However, the ATCOs were not diligent filling the sleep log, and that led to missing data.

It worth noting that Group 4 showed a lower level of boredom and mental workload compared to the other groups. Although this s a handful of ATCOs they are nevertheless representative of the operational staff currently employed at that small airport, comprising a total of 8/10 operators. The main difference between this sub-sample and the other two is the rostering system articulated on only two shifts excluding service on nights.

7. General Discussions and Conclusions

Shift work, a defining feature of air traffic controllers' lives, exposes them to diverse shift patterns and potential fatigue risks. The two case studies reported here investigated fatigue through an agile and user-friendly “toolkit” measuring sleep duration and quality, mental workload, boredom, and vigilance. This approach allows systematic data collection across various operational scenarios without disrupting existing rostering systems.

The observed homogeneity across shifts in workload and boredom is encouraging, suggesting that the current rostering system may generate little difference in these areas. However, vigilance shows a different story, with a slight reduction in the night shift, warranting further investigation.

Building upon established methods, this study incorporates the Psychomotor Vigilance Task (PVT). This validated test measure’s reaction time to visual cues, serving as a sensitive indicator of sleepiness, fatigue, and sleep disruption. Implementing the PVT alongside the subjective measures provides a multifaceted approach to understanding fatigue risk in ATC. The PVT has been successfully used in a great variety of settings and Detrimental effects of the night shift on vigilance have been reported for many types of shift workers, including nurses [10]. More inherent to the area we are discussing, Peukert and Meyer [11] used the PVT to investigate fatigue in air traffic controllers, revealing that fatigue levels varied across shifts and over time. Their results identified strategies to mitigate excessive vigilance decline and highlighted the impact of scheduling on subsequent fatigue levels. They even suggested optimal inter-shift rest periods for improved vigilance.

The most critical limitation shown by these case studies remains the willingness of operators to fill out the questionnaires and perform the PVT test diligently. Unfortunately, the attempt to use a liberal strategy to get operators to self-regulate was found to be suboptimal.

What is the overall lesson learned regarding the procedural aspects? Primarily, it would be beneficial to introduce specific training to educate on self-monitoring the ATCOs. That would enormously help ongoing data collection, which is jeopardized by the “indolence” of the operators. For example, despite taking less than a minute to complete, the sleep log journal has not entered the habits of most ATCOs. Also, collecting data through each controller's personal device would be helpful so that everyone can receive tests at the beginning and end of the shift, reducing the chance of forgetting about it. Finally, ATCOs' efforts could be rewarded to some extent (e.g., an extra day off) to increase compliance. Getting a benefit could mitigate the somewhat superficial attitude of some participants.

In any case, despite the limitations, the information obtained from these two case studies is valuable for developing minimally invasive fatigue monitoring that provides systematic information on the functional status of operators.

Author Contributions: F.D.N. and M.T. conceptualized the study; F.D.N, M.T, and G.T. managed and supervised data collection; F.D.N and G.T. wrote the initial draft. F.D.N. run the data analysis. All authors contributed to the interpretation of the results and manuscript editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Deep Blue Srl with a research contract to F.D.N.

Data Availability Statement: data are not available due to ethical or privacy restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Botella, J., Contreras, M. J., Shih, P. C., & Rubio, V. Two short tests fail to detect vigilance decrements. *European Journal of Psychological Assessment*. **2001**, 17(1), 48.
2. EUROCONTROL. Fatigue and Sleep Management: Personal Strategies for Decreasing the Effects of Fatigue in Air Traffic Control, **2007** (March). Brussels, Belgium.
3. Terenzi, M., Ricciardi, O., & Di Nocera, F. Rostering in air traffic control: a narrative review. *International journal of environmental research and public health*. **2022,** 19(8), 4625.
4. Knauth, P. Extended work periods. Industrial health. **2007**, 45(1), 125-136.

 doi:10.20944/preprints202405.1120.v1

5. Matthews, G., Warm, J. S., Reinerman, L. E., Langheim, L. K., & Saxby, D. J. Task engagement, attention, and executive control. *Handbook of individual differences in cognition: Attention, memory, and executive control.* **2010**, 205-230.
6. Holding, D.H. Fatigue, in G.R.J. Hockey (ed.), Stress and Fatigue in Human Performance. **1983**, Chichester, UK: Wiley.
7. Dinges, D. F., & Powell, J. W. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior research methods, instruments, & computers.* **1985**, 17(6), 652-655.
8. Hart, S. G., & Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *In Advances in psychology*, **1988,** 52, 139-183.
9. Fahlman, S. A., Mercer-Lynn, K. B., Flora, D. B., & Eastwood, J. D. Development and validation of the multidimensional state boredom scale. Assessment, **2013**, 20(1), 68-85.
10. Çolak, M., & Esin, M. N. Factors affecting the psychomotor vigilance of nurses working night shift. International Nursing Review. **2024**, 71(1), 84-93.
11. Peukert, M., & Meyer, L. The influence of rostering factors on fatigue in an irregular shift system in air traffic control. In 10th International Conference on Research in Air Transportation (ICRAT'22), Florida (USA), June 19-23, **2022**.