Physiological and Biochemical Measures of Mental Workload of Air Traffic Controllers: A Systematic Literature Review

Wiediartini, Udisubakti Ciptomulyono, and Ratna Sari Dewi
Industrial and Systems Engineering Department
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
wiwid@ppns.ac.id, udisubakti@ie.its.ac.id, ratna@ie.its.ac.id

Abstract

Air Traffic Control (ATC) is a complex cognitive task. The complexity may cause the workload of air traffic controllers (ATCs) to increase. Increased workload could degrade operator performance, which further affect safety and functioning of the system. Numerous studies in evaluating mental workload in various tasks used physiological and biochemical measures, including studies in ATCs. The use of physiological measures in those studies provides unique information of the operator's condition. This systematic review summarizes literatures on the measurement of the mental workload in ATCs using physiological and biochemical measures. The conducted systematic search come up with thirty-four studies to include for analysis. Physiological measures are categorized into cardiovascular, ocular, brain, respiration, and skin measures. Biochemical measures consisted of cortisol and salivary immunoglobulin (sIgA). The literature review covers various studies either in simulation environment or real working environment, all with different conditions and task scenarios. Even though this review specifically focuses on mental workload of ATCs, in general, the result of physiological measures still differs between broad studies. The difference might be influenced by the study task loads, difficulty levels, and the participants' characteristics.

Keywords

Air Traffic Control, Physiological Measures, Biochemical Measures, Mental Workload, Systematic Review.

1. Introduction

The nature of work has changed from physical work to demanding cognitive work as the increasing use of Visual Display Terminal (VDT) and automation system in many workplaces. The Air Traffic Control is a complex (Nolan 2011) and stressful job (Nikolova and Danev 2011).

The increase in traffic density and system automation become a challenge for air traffic controllers. The air traffic controllers should be able to sustain their focus of attention over a continuous period of time (Warm et al. 2008). As safety becomes the major priority of this task, the performance of air traffic controllers should persist. Therefore, mental workload becomes an important variable to investigate user performance (Young et al. 2014). Mental workload is often described with terms such as mental strain and emotional strain (Cain 2007). A repetitive simple task may not be affecting mental workload, but additional temporal pressure may affect mental workload (Young et al. 2014).

Workload is assessed using three general categories i.e., subjective measure, physiological measure, and performance measure. The development of the new technologies in physiological measure encourage numerous researches in mental workload. This type of measurement attempts to associate physiological changes with levels of workload. It is a natural type of workload index (Young et al. 2014). One of the advantages of physiological measures is the ability to provide real-time data. It also offers more accurate reports of mental workload, can be standardized, and compared across different studies (Tao et al. 2019). The increase in mental workload requires more cognitive resources, and it will provoke several physiological activities in the human body including cardiac activities, brain's electrical activities, eye movement, and metabolic changes (Fairclough and Houston 2004).

Air traffic control tasks required more mental resources than physical since it does not only include visual and auditory monitoring but also decision making. Most of the researchers manipulate the complexity of the task, as an indicator of workload level, by changing the number of aircraft. Some other factors that affect workload of ATCs are the number of potential conflicts, climbing/descending aircraft, and hand-offs, heading and speed differences, and presence of weather (Parimal 2007).

A review of mental workload using physiological measures has been conducted by (Tao et al. 2019; Kramer 1990; Lean and Shan 2011; Charles and Nixon 2019) which covered various domains of tasks. Marquart et al.(2015) reviewed more specific physiological measures (eye measure) in driving and found consistent results in pupil diameter, blink duration, and blink latency to changes in mental workload. The purpose of this review is to summarize the result of the previous studies that have been conducted which specifically concerns in physiological and biochemical measure of mental workload in air traffic control.

2. Method

A systematic literature search was conducted using database of Scopus published from the initiation of database until 20 February 2020. Keywords used for this search were ‘air traffic control’ AND ‘mental workload’ OR ‘cognitive’. The articles selected in this review are limited to full-text publication which is written in English. Duplicate studies that use the same sample information were only included the ones that analyse more physiological measures.

The title and abstract of the articles found from the search were read to verify if it meets the criteria. Only studies related to mental workload of air traffic controllers that includes one or more physiological and biochemical measure are included in this review. Afterward, the full text of potentially relevant studies were read to confirm its suitability for the review. Relevant articles from reference list of the included studies were manually searched to avoid any missed articles.

The physiological and biochemical measures were grouped based on the measurement used by the studies included in this review and also referred to previous studies by (Tao et al. 2019; Lean and Shan 2011; Charles and Nixon 2019). The physiological measures were grouped into cardiovascular, ocular, brain, respiration, and skin. Whereas the biochemical measures were cortisol and salivary immunoglobulin (sIgA).

3. Result

A literature search from Scopus resulted in 686 articles, the range of publication year start from 1978 to 2020. After verifying the title, abstract, and full text, 657 studies were excluded. Twenty-nine studies were included for review and 5 studies from manual search were added. A total of 34 studies were included for review (Figure 1). Data are summarized in Table Appendix A. Age, years of experience, gender, and type of participants were not reported in all studies.

3.1. Study Characteristics

The studies were conducted in real working situations (15%) and simulated air traffic control (85%). The sample size varied from 1 to 205 people and recruited professional air traffic controllers (ATCs), ATC students, university students, and other participants. The percentage of men ranged from 15% to 100%. Twelve studies did not report the gender of the participants. Two studies investigated the differences between expertise of the participants (Bernhardt et al. 2019; Hyun et al. 2006).

Table Appendix B summarizes the characteristic of studies included in this review. Sixty-four percent (64%) of the studies were conducted in this last decade. The majority of the studies used brain as physiological measure of mental workload (35%), followed by ocular (26%), cardiovascular (24%), respiration (7%), and skin (2%). Another 6% of the studies used cortisol (4%) and sIgA (2%). Most of the studies only used one group of physiological or biochemical (76%), 2 groups (15%), 3 groups (3%), and 4 groups (6%).

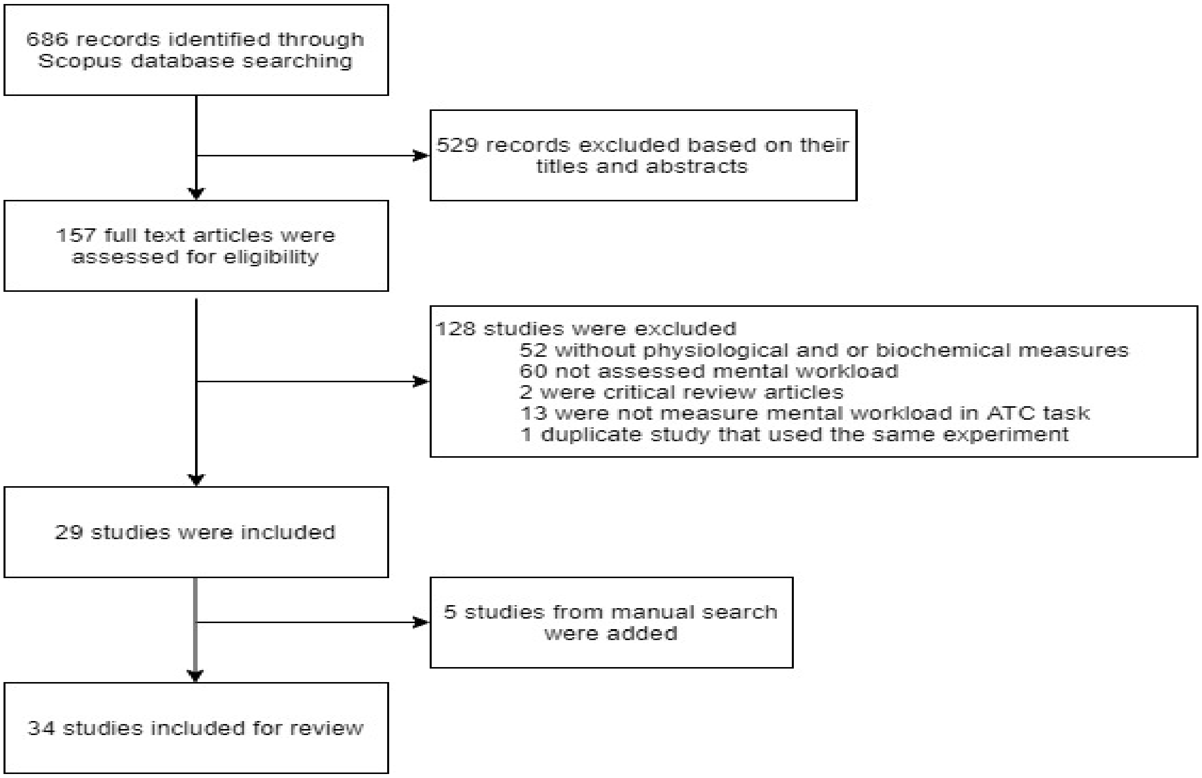

Figure 1. Procedure of study selection.

3.2. Physiological measure

3.2.1. Cardiovascular Measure

Most of the studies that measure cardiac activity used ECG techniques i.e., Heart Rate and Heart Rate Variability. Heart rate (HR) is the number of heart beats per minute whereas heart rate variability (HRV) represents variation of beat to beat interval. Measurement of heart rate variability can be divided into time-domain method, frequency-domain method, and non-linear method (Lean and Shan 2011; Laborde et al. 2017). Time-domain method as the simplest method measure both heart rate and the interval between successive normal complexes. Four recommended measurements for time-domain are SDNN, HRV triangular index, SDANN, and RMSSD (pNN50 and NN50) (Malik 1996). The frequency-domain can be conducted in short-term recording (2-5 minutes) and long-term recording (24 hours). Three components are categorized from short term recording: very low frequency (VLF; 0.0033-0.04 Hz), low frequency (LF; 0.04-0.15Hz), and high frequency (HF; 0.15-0.40 Hz). The long-term recording measure ultra-low frequencies (ULF; <0.0033 Hz) (Malik 1996; Forte et al. 2019).

Raduntz et al. (2019) evaluated heart rate and inter-beat interval of 21 participants while performing scenarios in interactive real-time ATC simulator. The result showed that LF decreased and HF increased with increasing load. On the contrary, Metzger and Parasuraman (2001) that recruited 18 professional ATCs found a trend of lower LF with higher mental workload. The nerves that affected LF is still controversial. Lean and Shan (2011) mentioned that LF represents mainly cardiac sympathetic, while Reyes et al. (2013) proposed that LF is mainly determined by parasympathetic system. Therefore, in their studies, Lean and Shan (2011) showed that as workload increase (sympathetic tone increase), LF increased, HF decreased, and LF/HF increased. High frequency reflects vagal tone (Laborde et al. 2017; Malik 1996), and LF/HF ratio indicates the balance of sympathetic and vagal tone (Lean and Shan 2011). High and medium workload levels were found to elicit significant reciprocally coupled sympathetic activation and parasympathetic inhibition (Backs et al. 2000). Similar result reported by Nikolova and Danev (2011) that the effect of mental workload might decrease parasympathetic activity and increase sympathetic activity. Heart rate is a result of contribution from sympathetic and parasympathetic inputs that have opposing effects (Backs 2000).

Observation of heart rate and blood pressure of 40 air traffic controllers during the working day was conducted by (Ereminas 2009). It found that the heart rate increases due to an increase in mental and physical work, and heart rate rhythm return back to normal as physical stress removed. HR found to increase with increasing task load (Raduntz et al. 2019; Back et al. 2000; Lille and Burnod 1983; Rose and Fogg 1993). However, no significant differences in HR and HRV are found between scenarios (Raduntz et al. 2019; Metzger and Parasuraman 2001; Brookings et al. 1996; Kabe et al. 2007). In contrast, Backs et al. (2000) that used the same scenario as Brookings et al (1996), found significant differences between level of workload. This contrary result can be caused by different expertise participants used in that study. Brookings et al (1996) used professional ATC specialists, while Backs et al (2000) used university students as the participants.

The findings come to different suggestions from the researchers. Brookings et al. (1996) suggested that both heart rate and heart rate variability were not sensitive to the different levels of mental workload due to the lack of significant changes. Nikolova and Danev (2011) also stated that heart rate is less sensitive measure of mental workload than HRV. However, Raduntz et al. (2019) and Kaber et al. (2007) found that HR measure is more diagnostic than HRV.

Another cardiovascular measure is blood pressure measures. Only one study measures systolic and diastolic arterial blood pressure during ATCs working day. The result showed a relationship between age and systolic blood pressure and found increase in participants aged 45-55. Blood pressure varied among the experiences of the operators and diastolic rises more than systolic blood pressure (Ereminas 2009).

3.2.2. Respiration

Respiration rate is the number of breaths take per minute. Respiration was found to be faster during high and medium load compared to the low load (Backs et al. 2000; Brookings et al. 1996). It was also found that respiration rate was significantly faster than the baseline in all three scenarios. Both of the findings come to the same conclusion although they use different levels of experiences among the participants.

3.2.3. Skin measures

Only one study included in this review used skin measure to observe workload of ATCs. Averty et al. (2002) used five physiological parameters: skin potential, skin conductance, skin blood flow, superficial skin temperature, and Instantaneous Heart Rate. Observation was conducted in real work environment with twenty-five participants of qualified air traffic controllers. Skin conductance and Instantaneous Heart Rate found to be the most correlated factors with three workload indexes (Traffic Load Index, Number of aircraft-N, NASA-TLX). Another three parameters (skin potential, skin blood flow, superficial skin temperature) found to be less reliable.

3.2.4. Eye measure

Parameters of eye movement used by the researchers to measure ATCs mental workload were blink rate, blink duration, number of fixations, fixation duration, total number of saccades, saccadic amplitude, saccadic peak velocity, saccadic duration, pupil diameter, and dwell time.

Ahlstrom and Friedman (2006) investigated the workload of ATCs from the use of weather displays. It reported that blink duration was significantly shorter in higher workload level and blink rate did not find to be decreasing with an increasing task load. The result of the blink rate observation was against the finding of Brookings et al. (1996) that found the decrease of blink rate as the difficulty of the task increases. Both of the studies recruited ATCs operator but has different range of age and length of experience. The participant used by Brookings et al. (1996) were ranged from 21 to 29 years of age with 2.5-7.5 years experience, whereas the median age of participants in a study of Ahlstrom and Friedman (2006) was 36.8 years and median experience were 13 years. Review by Kramer (1990) also found different research results related to blink rate, which may be caused by visual demand. Therefore, he suggested that more research is needed before blink rate could be used to measure mental workload.

Pupil diameter is correlated to workload Rodriguez et al.(2015) and found to be larger in higher workload levels (Bernhardt et al. 2019; Ahlstrom and Friedman 2006; Martin et al. 2011; Truschzinski 2017). Pupil diameter differed significantly between scenarios Martin et al. (2011); Truschzinski (2017) but not between expertise (Bernhardt et al. 2019). Pupil diameter was more sensitive to workload from visual rather than EEG workload metric (Bernhardt et al. 2019). This finding is similar to the result of Mandrick (2016) who observed the participants performed the Toulouse n-back task. Marchitto et al. (2016) investigated the effect of complexity on mental workload by dividing the task into conflict and no conflict. In conflict, large saccades showed less burst power with higher task complexity.

Peak velocity and duration of large saccades decreased whereas duration of micro saccades increased in higher cognitive complexity. Both large saccades and micro saccades differed significantly in conflict and in no conflict situations. During a simulated ATC task involving a decision task, Di Stasi et al. (2010) also found a decrease in saccadic peak velocity as cognitive load increase but found no effect of task complexity to saccade duration and velocity. The same result was reported by Brookings (1996) that saccade measures did not demonstrate significant differences across workload conditions.

Marchitto et al. (2016) found that the total number of fixations and mean fixation duration were not significant main effects of cognitive complexity, but fixation duration was correlated with attentional and cognitive effort in visual tasks. More complex tasks resulted in more frequent fixations. Hyun et al. (2006) investigated the number of fixations and the duration of each fixation between expert and novice air traffic controllers. It found that number of fixations and the duration of each fixation were not affected by controller expertise or complexity of traffic flow. The data showed that number of fixations was smaller in complex than in simple traffic flow for experts. On the other hand, number of fixations was greater in complex traffic compared to simple traffic flow for novices. Fixations and dwells are less and shorter in high traffic mix predictability with automation support (Rovira and Parasuraman 2006). The type of aircraft also affects the number of fixation (Martin et al. 2011).

3.2.5. Brain

The development of the technologies used for brain-based measures has increased the variety of research using this parameter. Functional near-infrared (fNIR) is a wearable device that can be used during the real task of the operators to measure cerebral hemodynamic responses (Ferrari and Quaresima 2012). The fNIR spectroscopy is used to measure concentration changes in blood oxygenation (Ayaz et al. 2012). Studies of mental workload for air traffic controller using fNIR technology has been conducted by (Ayaz et al. 2012; Harrison et al. 2013; Harrison et al. 2014). As measured by fNIR, the average oxygenation level is increased as the number of the aircraft count increases (Ayaz et al. 2012; Harrison et al. 2013; Ayaz et al. 2011). In 2014, Harrison et al. (2014) investigated the advantages of cognitive workload measure using fNIR compared to subjective workload assessment keypad. It was found that cognitive workload of air traffic controller increased faster using fNIR rather than the perceived workload obtained through workload assessment keypad. From all of these findings, it should be noted that not all of brain areas involved in air traffic control could be measured using fNIR technology (Ayaz et al. 2011).

The studies that observed between experience groups have been conducted by (Bernhardt et al. 2019; Ayaz et al. 2011). During complex cognitive and visuomotor tasks, haemoglobin concentration differed between beginner, intermediate and advanced levels. The higher the expertise, the lower the haemoglobin concentration found (Ayaz et al. 2012). However, Bernhardt et al. (2019) found no significant differences in EEG workload metric between novices and experts although the EEG workload metric was sensitive to workload scenarios.

Most of the studies analysed theta and alpha brain rhythms obtained from prefrontal cortex (PFC) and the posterior parietal cortex (PPC) regions to measure mental workload (Arico et al. 2016). Shou and Ding (2013) suggested frontal theta EEG as a sensitive and reliable metric to assess mental workload, due to the significant positive correlation between rhythmic theta activities and workload level during an ATC task. There is a significantly increased theta band at the high workload level between session and subject (Shou and Ding 2013) .

Dasari et al. (2017) analysed potential indicators of mental fatigue caused by the time on task using continuous EEG data. High cognitive workload done by air traffic controllers has induced mental fatigue after a certain time on tasks. The observation of the patterns in means and variances of EEG activity within theta, alpha, and beta indicates transitions among mental states. The transition could be trigger by mental fatigue which was shown approximately 70 minutes after the time on task.

The machine learning approach to estimate mental workload of ATCs has been used by some researchers. Arico et al.(2015) evaluated the mental workload of twelve ATC trainees based on brain measure using machine learning and showed significant differences of mental workload across three scenarios. Another study by Arico et al. (2016) investigated the use of passive brain-computer interface to evaluate mental workload. It was reported that EEG-based has high significant correlations to subjective workload measures. This study also compared SWLDA and the asSWLDA (automatic-stop stepwise linear discriminant analysis) algorithms to compute EEG-based mental workload indexes (WEEG). It found that asSWLDA was able to reduce the number of brain features and EEG channels. Arico et al. (2016) proposed a passive Brain-Computer Interface (pBCI) system to trigger Adaptive Automation (AA) solution in Air Traffic Management. The pBCI enables them to improve the human-machine interaction (HMI) by using mental workload as measurement using the covert brain activity. EEG signal was used to compute EEG-based mental workload index through asSWLDA. It was reported that when Adaptive Automation is activated, EEG-based mental workload index decreased. Subsequently, EEG-based mental workload index in hard tasks was not significantly higher than the easy task. Wilson and Russell (2003) used data from Brookings et al. (1996) to investigate if psychophysiological data would be able to be used to classify different mental workload using machine learning (ANN and SWDA). All of the studies using machine learning conducted by Arico et al. 2015; Wilson and Russell (2003); Borghini et al. 2014; Sciaraffa et al. (2019) showed promising results to differentiate the level of mental workload using EEG signal.

3.3. Biochemical

Cortisol and immunoglobulin A (IgA) taken out from blood, salivary, or urinary sample sometimes used to evaluate mental workload (Lean and Shan 2011). Zeier (1994) investigated workload and stress of 205 air traffic controllers in Zurich and Geneva using saliva sample. Saliva sample was taken before and after a working session to assess the concentration of secreted cortisol. Objective workload was computed from a combined index (consisting of the average number of aircraft to be controlled, number of radio-communications, and duration of radio-communications measured during successive 10 min intervals). To capture high and low load tasks, February was chosen as a period with low traffic density and May as a period with high traffic density. Cortisol levels sample were found to be higher in February than in May for sample before working session. All participants showed a slight increase from before working session to after working session. Both in February and May, result of cortisol assays of total sample were significantly correlated to objective workload and difficulty of work.

Another study using saliva conducted in 1996 by (Zeir 1996). Saliva samples were taken from 158 male air traffic controllers before and after working sessions. The result showed that demanding work of air traffic controllers increases salivary cortisol and also sIgA. The effect of before and after working session were similar for both salivary cortisol and sIgA. The salivary cortisol response was significantly correlated with actual and perceived workload. Conversely, the increase in sIgA was not correlated with actual and perceived workload.

4. Discussion

This review identified the result of studies in mental workload, specifically in ATCs. The physiological measures used by the researchers were in these categories: cardiovascular, ocular, brain, respiration, and skin. The biochemical measurements used were cortisol and salivary immunoglobulin (sIgA). Most of the studies use brain measure (16 articles) and only one study use skin measure. The extensive use of brain and eye measures seems instinctive as air traffic tasks mostly demanded cognitive effort in visual tasks. The studies conducted in real working environment identified in this review used heart rate and skin (3 articles) and cortisol and sIgA (2 articles). Other studies were conducted in simulated air traffic control (29 articles) with various scenarios of workload.

Raduntz et al. (2019) observed heart rate and inter-beat interval and found significant workload differences only in heart rate parameter. Both LF and HF band were not significantly differed between workloads, so he concluded that heart rate is the best indicator for mental workload. Similar conclusion reported by Kaber et al. (2007) that HR measure is more diagnostic than HRV. Contrary, study conducted during working hours of air traffic controllers from Sofia airport reported that HR to be less sensitive than HRV (Nikolova and Danev 2011). While Brookings (1996) found that none of HR and HRV were sensitive to the different levels of mental workload. Participants of the studies conducted by Nikolova and Danev (2011); Brookings (1996) were professional air traffic controllers, whereas the studies conducted by Raduntz et al. (2019); Kaber et al. (2007) were not.

Although using the same scenarios, Backs et al. (2007); Kaber et al. (2007) found different results in significant differences between level of workload which might be caused by different participants used in those studies. Brookings (1996) recruited professional ATCs as participants and found no significant differences in HR and HRV, while Backs et al. (2000) recruited university students and found significant differences between level of workload. However, the result for respiration was similar between the finding of Backs et al. (2000); Brookings (1996) which is faster respiration in high load.

Skin conductance was the only parameter in skin measure reported by Averty et al. (2002) as the most correlated parameter with three workload indexes. Another three parameters (skin potential, skin blood flow, superficial skin temperature) were less reliable.

There were 10 parameters of eye movement measure identified in this review. As air traffic control tasks include visual monitoring, eye movement widely used to measure mental workload. Several studies that used pupillometry found consistent results. Pupil diameter is increased as workload increases and is sensitive to changes in mental workload (Bernhard et al. 2019; Ahlstrom and Friedman-Berg 2006; Rodríguez 2015; Martin et al. 2011; Truschzinski et al. 2017). Saccadic peak velocity decreased as workload increases (Marchitto 2016; Rovira and Parasuraman 2006). During the observation between expertise, it was found that there are no significant differences in pupil diameter Bernhardt (2019), number of fixations, and the duration of each fixation Hyun et al. (2006) between experts and novices. Conflicting result in blink rate was found in the study of (Brookings 1996; Ahlstrom and Friedman-Berg 2006). As the workload increased, blink rate showed a decrease Brookings (1996), but Ahlstrom and Friedman-Berg (2006) did not find a decreased blink rate.

Another type of widely used measure was EEG. All were reported to be sensitive to change in mental workload (Bernhardt 2016; Shou and Ding 2013; Dasari et al. 2017). The mental fatigue of the air traffic controllers while executing high cognitive workload was estimated to happen in 70 minutes after the time on task (Dasari et al. 2010). The result of the fNIR showed an increase in the average oxygenation level as the number of the aircraft count increased (Ferrari and Quaresima 2012; Harrison 2013; Ayaz 2011). A challenge to classify mental workload has been captured in studies using machine learning. The EEG signal was used to compute EEG-based mental workload index and it consistently showed high accuracy to differentiate level of mental workload (Arico 2015; Wilson and Russell 2003; Borghini 2014; Sciaraffa et al. 2019).

Biochemical evaluations are often used together with physiological measures, and it can be used to acquire data in real working situations. Cortisol and immunoglobulin A (IgA) were used to evaluate mental workload in the studies identified in this review. Both Zeier (1994); Zeier et al. (1996) which used biochemical measure, reported that salivary cortisol and sIgA were related significantly to workload (Zeier 1994; Zeier et al. 1996).

5. Conclusion

This review identified studies that used physiological measures in assessing mental workload of air traffic controllers. Even in the same task, the physiological measure did not remain consistent. For example, HR measure is more diagnostic than HRV Radüntz (2019); Kaber et al. (2007), but Nikolova and Danev (2011) reported otherwise. Blink rate found to decrease with increasing workload Brookings (1996), whereas Ahlstrom and Friedman-Berg (2006) did not find a decreased blink rate. The different results might come from different scenarios and participants used by the researchers. Length of the experiments Arico (2015), number of the participants Radüntz et al. (2019), experience levels Bernhardt et al. (2019); Backs et al. (2000), training Liu et al. (2017) demographic variables Tao et al. (2019) may cause different result between studies.

It is noticed that one of the differences in those studies was the place where the participants were observed. It is a challenge to compare real working situations and simulated tasks. The weakness of the simulated tasks is that the complexity and difficulty of the tasks were difficult to compare between studies. To present a task that is closely resembles real situation, some studies have been conducted in a realistic air traffic control task. The development of the device which is comfortable to wear with minimal disturbance is still needed.

Most of the physiological measures were reported to be sensitive to change in mental workload, but the different characteristics of the participants recruited in the studies may affect the results. The same task load can differentiate mental workload among operators Xie and Salvendy (2000) and might be influenced by factors such as time constraints, environment, or experience (Megaw 2005; Wickens 2008; De Waard 1996). As it was found that there is a relationship between age and arterial systolic blood pressure Ereminas (2009), the age of the participants should also be considered while recruiting the participants. The use of several physiological measures are recommended since the result of only one physiological measure may not as conclusive.

References

Ahlstrom, U., and Friedman-Berg, F.J., Using eye movement activity as a correlate of cognitive workload, *International Journal Industrial Ergonomics*, vol. 36, no. 7, pp. 623-636, 2006.

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., and Golfetti, A et al., Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment, *Frontiers in Human Neuroscience*, 2016.

Arico, P., Borghini, G., Di Flumeri, G., Colosimo, A., Graziani, I., and Imbert, J-P et al., Reliability over time of EEG-based mental workload evaluation during Air Traffic Management (ATM) tasks, *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015.

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., and Babiloni, F., A passive brain–computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks, *Lab Experiments to Real-World Applications*, vol. 228, pp. 295-328, 2016.

Averty, P., Athenes, S., Collet, C., and Dittmar, A., Evaluating a new index of mental workload in real ATC situation using psychophysiological measures, Proceedings The 21st Digital Avionics Systems Conference. *IEEE*, vol. 2, 2002.

Ayaz, H., Shewokis, P.A., Bunce, S., Izzetoglu, K.., Willems, B., and Onaral, B., Optical brain monitoring for operator training and mental workload assessment, *NeuroImage*, vol.59, no. 1, pp. 36-47, 2012.

Ayaz, H., Willems, B., Bunce, S., Shewokis, P.A., Izzetoglu, K.., and Hah, S., et al, *Estimation of Cognitive Workload during Simulated Air Traffic Control Using Optical Brain Imaging Sensors*, Lecture Notes in Computer Science, pp. 549-558, 2011.

Backs, R.W., Navidzadeh, H.T., and Xu, X., Cardiorespiratory Indices of Mental Workload during Simulated Air Traffic Control. *Proceedings of the Human Factors and Ergonomics Society*, vol. 44, no. 13, pp. 89-92, 2000.

Bernhardt, K.A., Poltavski, D., Petros, T., Ferraro, F.R., Jorgenson, T., and Carlson, C., The effects of dynamic workload and experience on commercially available EEG cognitive state metrics in a high-fidelity air traffic control environment, *Applied Ergonomics*, vol. 77, pp. 83-91, 2019.

Borghini, G et al., *Analysis of neurophysiological signals for the training and mental workload assessment of ATCos*, in SIDs 2014 - Proceedings of the SESAR Innovation Days, 2014.

Brookings, J.B., Wilson, G.F., and Swain, C.R., Psychophysiological responses to changes in workload during simulated air traffic control, *Biological Psychology*, vol. 42, no. 3, pp. 362-377, 1996.

Cain, B., *A Review of the Mental Workload Literature*, Defence Research and Development Canada Toronto, 2007.

Charles, R.L., and Nixon, J., Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, vol. 74, pp. 221-232, 2019.

Dasari, D., Crowe, C., Ling, C., Zhu, M., and Ding, L., EEG pattern analysis for physiological indicators of mental fatigue in simulated air traffic control tasks, *PsycEXTRA Dataset*, vol. 1, no. 3, pp. 205-209, 2010.

Dasari, D., Shou, G., and Ding, L., ICA-Derived EEG Correlates to Mental Fatigue, Effort, and Workload in a Realistically Simulated Air Traffic Control Task, *Frontiers in Neuroscience*, vol. 11, 2017.

De Waard, D., *The Measurement of Drivers ' Mental Workload*, The Traffic Research Centre, University of Groningen, Haren, Den Haag, 1996.

Di Stasi, L.L., Marchitto, M., Antolí, A., Baccino, T., and Cañas, J.J., Approximation of on-line mental workload index in ATC simulated multitasks, *Journal of Air Transport Management*, vol. 16, no.6, pp. 330-333, 2010.

Ereminas, D., An Observational Workplace Study Of Cardiovascular Variables In Air Traffic Controllers, *Aviation*, vol. 13, no. 2, pp. 50-55, 2009.

Fairclough, S.H., and Houston, K..A., Metabolic measure of mental effort. *Biological Psychology*, vol. 66, no. 2, pp. 177-190, 2004.

Ferrari, M., and Quaresima, V., A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application, *Neuroimage*, vol. 63, no. 2, pp. 921-935, 2012.

Forte, G., Favieri, F., and Casagrande, M., Heart Rate Variability and Cognitive Function: A Systematic Review, *Frontiers in Neuroscience*, vol, 13, pp. 710, 2019.

Harrison, J., et al., *Human performance assessment study in aviation using functional near infrared spectroscopy*, in Foundations of Augmented Cognition, vol. 8027, pp. 433-442, 2013.

Harrison, J., Izzetoglu, K., Ayaz, H., Willems, B., Hah, S., and Ahlstrom, U., et al, Cognitive Workload and Learning Assessment During the Implementation of a Next-Generation Air Traffic Control Technology Using Functional Near-Infrared Spectroscopy. *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 4, pp. 429-440, 2014.

Hyun, S.H., Kim, K.T., Song, SM., Yoon, J.S., Lee, J.I., and Cho, S.M., Effects of expertise and situation complexity on visual attention and action planning for air traffic control, *Human Factors and Ergonomics Society Annual*, vol. 50, no. 3, pp. 420-423, 2006.

Kaber, D.B., Perry, C.M., Segall, N., and Sheik-Nainar, M.A., Workload State Classification With Automation During Simulated Air Traffic Control, *International Journal of Aviation Psychology*, vol. 17, no. 4, pp. 371-390, 2007.

Kopardekar, P., *Airspace Complexity Measurement : An Air Traffic Control Simulation Analysis*, in 7th USA/Europe Air Traffic Management R/D Seminar, 2007.

Kramer, A.F., *Physiological Metrics of Mental Workload: A Review of Recent Progress,* Defense Technical Information Center, 1990.

Laborde. S., Mosley, E., and Thayer, J.F., Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research – Recommendations for Experiment Planning, Data Analysis, and Data Reporting, *Frontiers in Psychology,* vol. 8, pp. 213, 2017.

Lean, Y., and Shan, F., Brief review on physiological and biochemical evaluations of human mental workload. *Human Factors* and Ergonomics in *Manufacturing* & *Service*, 2011.

Lille, F., and Burnod, Y., Professional Activity and Physiological Rhythms1, *Advances in Biological Psychiatry,* vol. 11, pp. 64–71, 1983.

LIU, Y. et al., EEG- Based Mental Workload Recognition in Human Factors Evaluation of Future Air Traffic Control Systems, in Transdisciplinary Engineering: A Paradigm Shift, Proceedings of the 24th ISPE Inc, *International Conference on Transdisciplinary Engineering*, 2017.

Malik, M., Heart Rate Variability, *Annals Noninvasive Electrocardiol*, vol. 1, no. 2, pp. 151-181, 1996.

Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., and Causse, M., Neural and psychophysiological correlates of human performance under stress and high mental workload, *Biological Psychology,* vol. 121, pp. 62-73, 2016.

Marchitto, M., Benedetto, S., Baccino, T., and Cañas, J.J., Air traffic control: Ocular metrics reflect cognitive complexity, *International Journal of Industrial Ergonomics*, vol. 54, pp. 120-130, 2016.

Marquart, G., Cabrall, C., and Winter, D.J., Review of Eye-related Measures of Drivers' Mental Workload. *Procedia Manufactur ,* vol. 3, pp. 2854-2861, 2015.

Martin, C., Cegarra, J., and Averty, P., Analysis of Mental Workload during En-route Air Traffic Control Task Execution Based on Eye-Tracking Technique, *Lecture Notes in Computer Science*, pp. 592-597, 2011.

Megaw, T., *The definition and measurement of mental workload. Evaluation of Human Work, 3rd Edition*. CRC Press, pp. 525–551, 2005

Metzger, U., and Parasuraman, R., The Role of the Air Traffic Controller in Future Air Traffic Management: An Empirical Study of Active Control versus Passive Monitoring, *Journal Human Factor and Ergonomics Society*, vol. 43, no. 4, pp. 519-528, 2001.

Muñoz-de-Escalona, E., and Cañas, J.J., *Latency Differences Between Mental Workload Measures in Detecting Workload Changes,* Human Mental Workload: Model and Application, pp. 131–146, 2019.

Nikolova., and Danev, S., Effect of mental workload on autonomic cardiovascular control, *Acta Medica Bulgarica*, vol. 38, no. 2, pp. 15–21, 2011.

Nolan, M.S., *Fundamentals of Air Traffic Control, 5th ed.* NY, USA: Delmar, 2011.

Radüntz, T., Mühlhausen, N., Fürstenau, E., Cheladze, B., and Meffert, Cardiovascular Parameters for Mental Workload Detection of Air Traffic Controllers, *Intelligent Human Systems Integration 2019*, vol. 903, pp. 789-794, 2019.

Reyes., del Paso, G.A., Langewitz, W., Mulder, L.J.M., Roon, A., and Duschek, S., The utility of low frequency heart rate variability as an index of sympathetic cardiac tone: A review with emphasis on a reanalysis of previous studies. *Psychophysiology*, vol. 50, no. 5, pp. 477-487, 2013.

Rodríguez, S., Sánchez, L., López, P., and Cañas, J.J., *Pupillometry to assess Air Traffic Controller workload through the Mental Workload Model*, Proceedings of the 5th International Conference on Application and Theory of Automation in Command and Control Systems - ATACCS '15, 2015.

Rose, R.M., and Fogg, L.F., Definition of a responder: analysis of behavioral, cardiovascular, and endocrine responses to varied workload in air traffic controllers, *Psychosomatic Medicine*, vol. 55, no. 4, pp. 325-338, 1993.

Rovira, E., and Parasuraman, R., Controller performance and attention allocation in future air traffic management: Effects of pilot intent information, *PsycEXTRA Dataset*, 2006.

Sciaraffa, N., Aricò, P., Borghini, G., Flumeri, G.D., Florio, A.D., and Babiloni, F., On the Use of Machine Learning for EEG-Based Workload Assessment: Algorithms Comparison in a Realistic Task, *Human Mental Workload: Model Application*, pp. 170-185, 2019.

Shou, G., and Ding, L., Frontal theta EEG dynamics in a real-world air traffic control task, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5594-5597, 2013.

Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., and Zhang, T., A Systematic Review of Physiological Measures of Mental Workload. *International Journal of Environmental Research and Public Health*, vol. 16, no. 5, pp. 2716, 2019.

Trapsilawati, F. et al., Perceived and Physiological Mental Workload and Emotion Assessment in En Route ATC En Route ATC Environment: A Case Study, in Transdisciplinary Engineering: A Paradigm Shift, *International Conference on Transdisciplinary Engineering*, vol. 5, pp. 420–427, 2017.

Truschzinski, M., Betella, A., and Brunnett, G., Verschure PFMJ. Emotional and cognitive influences in air traffic controller tasks: An investigation using a virtual environment, *Applied Ergonomics,* vol. 69, pp. 1-9, 2018.

Warm, J.S., Parasuraman, R., and Matthews, G., Vigilance Requires Hard Mental Work and Is Stressful, *Journal Human Factor and Ergonomics Society*, vol. 50, no. 3, pp. 433-441, 2008.

Wickens, C.D., Multiple Resources and Mental Workload, *Journal Human Factors and Ergonomics Society,* vol. 50, no. 3, pp. 449-455, 2008.

Wilson, G.F., and Russell, C.A., Operator Functional State Classification Using Multiple Psychophysiological Features in an Air Traffic Control Task, *Journal Human Factors and Ergonomics Society*, vol. 45, no. 3, pp. 381-389, 2003.

Xie, B., and Salvendy, G., Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. *International Journal of Work, Health & Organisations*, vol. 14, no. 1, pp. 74-99, 2000.

Young, M.S., Brookhuis, K..A., and Wickens, C.D., Hancock PA. State of science: mental workload in ergonomics. *Ergonomics*, vol. 58, no. 1, pp. 1-17, 2014.

Zeier, H., Brauchli, P., and Joller-Jemelka, H.I., Effects of work demands on immunoglobulin A and cortisol in air traffic controllers, *Biological Psychology*, vol. 42, no. 3, pp. 413-423, 1996.

Zeier, H., Workload and psychophysiological stress reactions in air traffic controllers, *Ergonomics*, vol. 37, no. 3, pp, 525-539, 1994.

Appendix A

Table A. Summary of physiological and biochemical measures.

	Participant Characteristic					**Real work studies (R)/ Laboratory (L)**	**Physiological**					**Biochemical**	
Study	**N**	**Participants**	**Gender (% men)**	**Age M (SD)**	**Experience (year)**		**Cardio-vascular**	**Ocular**	**Brain**	**Respiration**	**Skin**	**Cortisol**	**sIgA**
(Nikolova and Danev 2011)	51	Professional ATC				R	v						
(Warm et al. 2008)	49	ATC students	95.9	20.97		L		v	v				
(Hyun et al. 2006)	17	Professional ATC	76.5			L		v					
(Radüntz et al.2019)	21		90.5	38 (11)		L	v						
(Metzger and Parasuraman 2001)	14	Professional ATC		36.4 (5.1)	12.4 (2.9)	L	v						
(Backs et al. 2000)	27		40.7	21		L	v			v			
(Ereminas 2009)	40	Professional ATC				R	v						
(Brookings et al. 1996)	8	Professional ATC	87.5	21-29	2.5-7.5	L	v	v	v	v			
(Kaber et al. 2007)	25			18-23		L	v						
(Averty et al. 2002)	25	Professional ATC	96			R	v				v		
(Ahlstrom and Friedman-Breg 2006)	6	Professional ATC		36.8 (6.3)	13 (6.7)	L		v					
(Rodríguez et al.2015)	1	Professional ATC				L		v					
(Martin et al. 2011)	37	Professional ATC	75.7	26-56	0.5-26	L		v					
(Harrison et al. 2013)	12	Professional ATC				L			v				
(Truschzinski et al. 2018)	25	University students	64	28.12 (5.67)		L		v					
(Marchito et al. 206)	26	University students	15.4	22 (2)		L		v					
(i=Di stasi et al. 2010)	23		39.1	24.6 (3.65)		L		v					
(Rovira and Parasuraman 2006)	16	Professional ATC		40.75 (6.07)	16.59 (4.96)	L		v					
(Ayaz et al.2012)	24	Professional ATC		24-55	3.0-30	L			v				
	7	University students		21-28		L							
(Ayaz et al. 2011)	24	Professional ATC		24-55		L			v				
(Aricò et al. 2016)	12	Professional ATC		40.4 (5.54)		L			v				
(Shou and Ding 2013)	11	University students	100	25 (4.3)		L			v				
(Dasari et al. 2017)	10		100	25 (4.3)		L			v				
(Dasari et al. 2010)	11		100	25 (4.3)		L			v				
(Arico et al. 2015)	12	ATC students		25 (3)		L			v				
(Aricò et al. 2016)	12	ATC students		23 (2)		L			v				
(Wilon and Russell 2003)	8	Professional ATC	87.5	21-29	2.5-7.5	L	v	v	v	v			
(Borghini et al. 2014)	6			21 (4)		L	v	v	v				
(Sciaraffa et al. 2019)	35	Professional ATC				L			v				

(Zeier 1994)	205	Professional ATC	97.1	42		R						v	
(Zeier et al. 1996)	199	Professional ATC	100	42 (6)		R						v	v
(Trapsilawati et al. 2017)	2	Professional ATC & Students	100	31 and 35		L			v				
(Liu et al. 2017)	36	Professional ATC & Students				L			v				
(Muñoz-de-Escalona et al. 2019)	32	University students	28.125	22.1 (22)		L		v					

Appendix B

Table B. Characteristics of the 33 studies included.

Characteristics	**N**	**%**		**Characteristics**	**N**	**%**
Year of publication				*Place of data collection*		
Before 2000	3	9%		Real working environment	5	15%
2000-2009	9	27%		Laboratory	29	85%
2010-2019	22	64%				
Measures				*Number of physiological or biochemical measures used*		
Cardiovascular	11	24%		1	26	76%
Ocular	13	26%		2	5	15%
Brain	16	35%		3	1	3%
Respiration	3	7%		4	2	6%
Skin	1	2%				
Cortisol	2	4%				
sIgA	1	2%				