Multimodal measurement of the mental workload during an assembly and disassembly task

Lorrys Berthon[a]*, Sylvain Fleury[b], Fabien Bernard[a], Raphael Paquin[a] and Simon Richir[b]

[a] *Airbus Helicopters, 13700 Marignane, France;*

[b] *Arts et Métiers, Sciences et Technologies, ENSAM Université, 53810 Changé, France*

lorrys.l.berthon@airbus.com

Lorrys Berthon is a PhD student and engineer specializing in Human Factors and Ergonomics at Airbus Helicopters and affiliated at ENSAM (Arts et Metiers Institute of Technology), with a particular interest in measuring mental workload in complex industrial environments. His work primarily focuses on the multimodal assessment of cognitive workload, applied to critical sectors such as aeronautical maintenance (helicopters) and virtual reality simulation environments.

Sylvain Fleury is a researcher and engineer at ENSAM, affiliated with the LAMPA laboratory. His work focuses on Human Factors, Ergonomics, and immersive technologies, exploring human-machine interaction and optimizing performance in complex systems, including virtual and augmented reality environments.

Fabien Bernard is a researcher and engineer at Airbus Helicopters and affiliated at ENSAM. Holding a Ph.D. and HDR, his work focuses on integrating Human Factors and Ergonomics into multidisciplinary settings. His research interests include human-machine interaction, maintainability in aeronautical systems, and the application of virtual reality simulation to optimize design, performance, and safety in industrial contexts.

Raphael Paquin is a researcher and engineer specializing in Human Factors and maintainability, particularly within the aeronautical maintenance sector (helicopters), often collaborating with Airbus Helicopters. His work focuses on integrating human factors into design and utilizing virtual and augmented reality to optimize maintenance tasks and human performance in complex operational environments.

Simon Richir is a prominent researcher and professor at ENSAM, specializing in virtual and augmented reality and human-computer interaction. His work focuses on the design and application of immersive technologies to enhance human performance, learning, and decision-making in complex systems.

Multimodal measurement of the mental workload during an assembly and disassembly task

Mental workload overload is a major cause of human error in industrial tasks such as maintenance. Human errors can compromise not only system safety but also lead to high social and economic costs, reduce equipment productivity, and cause incidents, accidents, and fatalities. To this day, we do not have an adequate assessment of mental workload in maintenance, which would help design maintenance processes more effectively by incorporating this crucial aspect. The objective of this study is to determine the ability of our indicators to measure mental workload during a disassembly and assembly task in a laboratory condition. Thirty-six participants performed a disassembly and assembly task under two different mental workload conditions. Subjective measures (NASA-TLX), performance metrics (number of errors), and cardiovascular data (heart rate, heart rate variability, and breathing rate) were analyzed. We observed a higher number of errors and elevated NASA-TLX scores in the high mental workload condition. Regarding cardiovascular data, interesting trends in the temporal domain were observed despite mostly non-significant results. Although conducted in a laboratory, this multimodal mental workload measurement method is promising for diagnosing and understanding operators' cognitive behavior, and deserves validation in real-world maintenance conditions.

Keywords: Mental workload, Human error, Heart rate variability, NASA-TLX, Maintenance.

I. Introduction

Maintenance is predominantly carried out by humans, which can pose a major threat to the safety of both operators and systems (Hobbs, 2021). The working conditions of maintenance technicians expose them to various challenges: postural and musculoskeletal constraints, carcinogenic chemicals, thermal and noise nuisances, vibrations, outdoor work exposed to the elements, as well as time pressure, and frequent interruptions (Grusenmeyer, 2023). In all fields, maintenance activities can expose operators to health and safety risks, but particularly in the aviation industry, where risks can be significantly more severe than in other sectors. In the aviation industry, among accidents attributed to human error, 20% are due to ineffective maintenance activities (Yiannakides and Sergiou, 2019). Human errors can have severe repercussions on flight safety, potentially leading to catastrophic outcomes (Hobbs and Williamson, 2002; Dhillon, 2009; Bernard et al., 2020; Nkosi, Gupta, and Mashinini, 2020). In maintenance, errors can remain latent for extended periods and thus represent an unexpected risk to the safety of a complex system. To optimize the quality and efficiency of maintenance, it is crucial to understand the skills and limitations of maintenance personnel during task execution. Sometimes, failures in maintenance quality go undetected for months or even years until they are discovered or lead to an operational incident (Hobbs, 2021).

A. Maintenance workload

To identify and understand the causes of human errors, it is necessary to comprehensively examine all contributing factors with the aim of improving system safety (Nkosi, Gupta, and Mashinini, 2020). The goal is to develop specific strategies and designs to significantly reduce the risk of human error. Analyzing operator behavior from a human factors perspective is the first step in designing safer systems. The daily aircraft maintenance activities, carried out under stressful conditions such as working at heights or accessing confined spaces, can induce excessive workload on operators, potentially compromising their performance and safety. Maintenance operators in the aviation industry face demanding environments that require both physical and mental capabilities, operating under consistent workload and pressure. Ensuring the health and safety of these operators is essential for the future of maintenance design (Bernard et al., 2020). To enhance maintenance quality and efficiency, it is crucial to understand the capabilities and constraints of maintenance operators while performing tasks. Operators' workload can be influenced by several interdependent factors such as task difficulty, time pressure, total working hours, rest periods, job nature, task sequence, and environmental conditions (Hobbs and Williamson, 2002; Yiannakides and Sergiou, 2019; Bernard et al, 2020; Hobbs, 2021). These interconnected factors can induce stress or fatigue, leading to reckless behaviors, poor performance, and errors by maintenance technicians, and can also compromise the situational awareness of maintenance operators (Santos and Melicio, 2019; Yiannakides and Sergiou, 2019). This is where the notion of physical and mental strain comes into play, i.e., the mental and physical ‘cost’ associated with the effects of this constraint on the individual. Workload is twofold: it can be physically demanding due to the efforts, postures, and loads handled by the operator, and mentally demanding due to the thinking, memorization, and planning required (Yiannakides and Sergiou, 2019; Santos and Melicio, 2019; Sugiharto, 2019; Nkosi, Gupta, and Mashinini, 2020). By assessing workload from the beginning of the maintainability design cycle, it would be possible to enhance maintenance activities to reduce human errors, through better integration of operators' capabilities. Anticipating this during design is the role of maintainability, defined as the ability of an element, under specific usage conditions, to be maintained or restored to a state in which it can perform a required function, provided maintenance is carried out under defined conditions (Dhillon, 1999; Zaki et al., 2019). To ensure the sustainability of a product or system, it is crucial to design a system that allows for efficient maintenance in various contexts, with minimal requirements for specialized skills and equipment (Bernard et al., 2020). Integrating Human Factors and Ergonomics (HFE) during the design phase helps anticipate and optimize interactions between operators and system components. Implementing a method to more effectively incorporate the physical aspects of HFE into maintainability has contributed to successful human-centered design (Bernard et al., 2023). However, the cognitive aspect should not be neglected, especially as little is known about the cognitive aspects of maintenance errors and the factors that contribute to them. Developing this approach marks an initial step towards incorporating cognitive HFE principles and their early integration into each phase of maintainability development.

B. Mental workload assessment

Measuring mental workload is crucial for developing new technologies, designing user interfaces and complex systems, and optimizing human-machine interactions, thereby enabling more precise identification of areas where users experience high levels of mental workload (Longo et al., 2022). Quantifying mental workload helps in predicting operator and system reactions more accurately. It promotes better task management and organization, thereby reducing human errors. In real operational situations, a significant gap between available resources and the workload imposed on the operator can lead to task disengagement or incidents related to work overload (Wickens, 2017). Therefore, it is essential to consider the complete operational environment in the research and application of new techniques for gathering information on mental workload (Young et al., 2014). Many methods aimed at assessing mental workload exist, but there is no universally accepted method. This is due to the variety of factors influencing mental workload and the different approaches that have been and continue to be developed in other specific disciplines (Leplat and Sperandio, 1967; Longo et al., 2022). Validating a multimodal model for measuring mental workload in connection with operational requirements would help predict performance failures in future systems. Assessing mental strain will aid in more precisely forecasting operator and system reactions. It is generally accepted that there are three main categories of measures for evaluating mental workload (Hart and Staveland, 1988; Martin, Hourlier, and Cegarra, 2013; Mandrick et al., 2013; Charles and Nixon, 2019; Le Gonidec, 2022; Kramer, 2020; Barajas-Bustillo et al., 2022): subjective measures, performance measures and physiological measures.

B.1. Subjective measures

Subjective measures typically involve a participant providing qualitative and/or quantitative feedback about their experience while performing a task (Charles and Nixon, 2019; Hart and Staveland, 1988; Le Gonidec, 2022). In most subjective measurement methods, the operator must complete the questionnaire after executing the task. Subjective assessments are the only source of information on the subjective perception of the impact of a task on operators, encompassing the influences of the many factors that contribute to workload. Among these different types of evaluations, multidimensional subjective assessments are the most commonly used for measuring mental workload. These measures provide diagnostic information on specific sources of load (mental, physical, temporal, and frustration), as well as a detailed overall summary. For example, the NASA-TLX combines information from its various criteria, reducing certain sources of variability among participants while highlighting the contributions of other significant sources of variability in the experimental context (Hart and Staveland, 1988). Retrospective subjective measures can introduce a memory bias and lead to a reinterpretation of sensations experienced as well as the operator's assessment of mental workload (Mallat et al., 2020; Le Gonidec, 2022) and also that the results are influenced by participant experience (Barajas-Bustillos et al., 2023).

B.2. Performance measures

Performance measures are used to assess the operator's ability to perform a task at an acceptable level and indirectly evaluate their mental workload in isolation. Additionally, they offer the opportunity to assess the quality of the task accomplished based on the objectives related to a particular task (Martin, Hourlier, and Cegarra, 2013). It is generally accepted that by optimizing mental workload operator performance can be improved (Cain, 2008; Young et al., 2014; Longo et al., 2022,). This technique allows for the collection of dependent variables such as response times, task execution durations, and error rates (Cegarra and Chevalier, 2008; Longo et al., 2022). Performance measures are divided into two main categories. The first, primary tasks, provide a direct and precise measure of performance, especially for long-duration mental workload. Their drawback is the inability to determine the cause of workload variations when several tasks are executed concurrently (Longo, 2015). The second category, secondary tasks, evaluates the operator's available capacity while they perform the primary task (Young et al., 2014). However, a disadvantage is their inability to highlight the source of variations in mental workload when multiple tasks are performed concurrently, which may lead to changes due to heightened vigilance from sensory interference caused by divided attention between primary and secondary tasks.

B.3. Physiological measures

Significant developments in the field of mental workload assessment have led to the emergence of various categories of physiological measures. These include cardiovascular, ocular, salivary, electrodermal, and neurophysiological measures (Young et al., 2014; Charles and Nixon, 2019; Longo et al., 2022). These physiological measures can reflect the actions of the autonomic nervous system, which comprises the sympathetic nervous system (SNS) that reacts to stress and stimulation, and the parasympathetic nervous system (PNS) that induces a general inhibition of the body’s functions. The role of these two systems is to maintain homeostasis, ensuring the stability of the body's internal state (Mandrick et al., 2013; Mallat et al., 2020; Le Gonidec, 2022). In situations of intense mental workload, the PNS decreases its activity while the SNS increases its activity to prepare the body for a physical or mental response. Conversely, when activation is not necessary, the SNS becomes less active, allowing the PNS to dominate and maintain basic bodily functions, thus promoting energy storage. Therefore, an increase in mental workload involves an increase in cognitive resources used by the individual, which is reflected in the physiological activity of the human body (Charles and Nixon, 2019; Tao et al., 2019; Kramer, 2020). However, the operator's movements (muscular, ocular, etc.) can introduce bias, and the multiplicity of factors to which physiological systems respond reduces their selectivity and diagnostic capability (Charles and Nixon, 2019).

C. *Quality criteria in mental workload measurement*

Cegarra and Chevalier (2008) point out that no single method can perfectly measure the mental workload level of an operator. Each approach should provide complementary information when combined with another. Combining task-performance and physiological measures can be problematic because performance metrics are often computed after task completion, while physiological signals are recorded continuously (Longo et al., 2022). The combination of physiological measures with subjective measures is the most common. However, from an experimental point of view and in the context of a field study, it is often difficult to implement a variety of measurement methods without interfering with the main task. Using multiple measures of workload allows for the acquisition of data across a broader spectrum because each measure has its own specificity (Xie and Salvendy, 2000; Cegarra and Chevalier, 2008). Therefore, to choose the appropriate combination of measures, consideration of the following criteria is necessary (Hart and Staveland, 1998; Cain, 2007; Cegarra and Chevalier, 2008; Longo et al., 2022):

- Sensitivity: Focuses on the ability to discern different levels of task demands. For example, physiological measures are generally sensitive indicators of mental workload.

- Selectivity: Requires the measure to remain constant when the workload remains constant. Unlike objective performance measures, physiological measures are generally not selective due to fluctuations caused by numerous external parameters such as inter-individual differences or signal noise.

- Diagnosticity: Focuses on the measure's ability to identify the source of the workload in the task. Subjective measures have relevant multidimensional diagnostic capacity. In particular, the NASA-TLX easily identifies sources of workload. However, heart rate measurement has slightly limited diagnostic capacity due to the difficulty in differentiating the causes of the workload.

In measuring the mental workload in a representative real-world environment, the combination of measurement methods is fundamental to understanding the origin of the load, its variations, and its limits. During maintenance, technicians often need to move actively and access areas that are physically and visually challenging. The frequent need to change positions and expend significant physical energy can affect the accuracy of physiological measurements collected for mental workload analysis (Berthon et al., 2024). An effective measure of mental workload requires a multimodal assessment of operator strain through the combination of multiple measurement requirements. In this context and based on the quality criteria outlined above, this study aims to validate a method for measuring mental workload during a disassembly and assembly task in a laboratory context, involving cognitive resources comparable to those required by a maintenance operator during their work. Cardiovascular data (heart rate, heart rate variability, and breathing rate) were used to determine the response to high mental workload. The NASA-TLX was used to diagnose the sources of workload for participants, while calculating the number of errors was used to determine the level of performance affected by the participants' mental workload. This paper examines a specific hypothesis for each category in our multimodal measure of mental workload, applied to a task similar to maintenance in a laboratory context. For the HMW group, we hypothesized that the number of disassembly and assembly errors would be higher. We also hypothesized that there would be a higher subjective perception of workload for HMW. Finally, we hypothesized that there would be a significant difference in cardiovascular data measures between groups.

II. Material & Methods

Thirty-six volunteers participated in this study. They were divided into two groups: High Mental Workload (HMW) and Low Mental Workload (LMW). The mean age in the HMW group was 29.5 ± 8.2 years, while in the LMW group it was 27.3 ± 5.7 years. Participants were not compensated for participation. In the HMW group comprised 4 women and 14 men; the LMW group comprised 3 women and 15 men. We allocated participants to the groups this way to balance the effects of sex on cardiovascular responses (Geovanini et al., 2020). None of the participants reported a history of cardiovascular disease. None had prior maintenance experience or had ever performed this task before. Participants completed a lifestyle questionnaire, which allowed us to assess their level of sedentariness because it can influence the cardiovascular activation level, and to identify exclusion criteria related to the consumption of psychoactive substances. The level of sedentariness among participants, calculated based on the amount of weekly physical activity including walking, was 6.3 ± 3.1 hours for the HMW group and 6.4 ± 2.9 hours for the LMW group. Regarding lifestyle habits, 100% of participants in HMW group did not use tobacco and 88.9% for LMW group. In the HMW group, 88.9% had not consumed alcohol in the previous 24 hours, compared to 83.3% in the LMW group. Additionally, 100% of the HMW group and 94.4% of the LMW group reported not using any psychoactive substances. Cardiac data from three participants were excluded: one due to psychoactive substance use and two due to data loss (processing error). Thus, cardiac analyses included 17 HMW and 16 LMW participants (out of 18 per group).

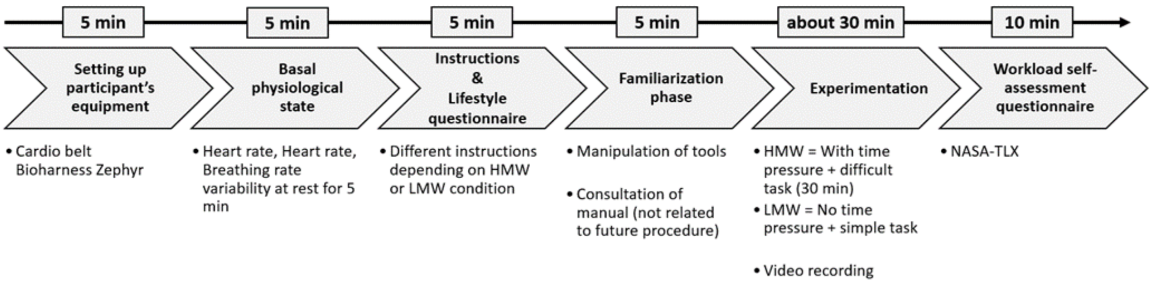


Figure 1. Protocol design

A. *Experimental protocol*

This study was approved by the Institutional Review Board under the following reference numbers: IRB00003888, IORG0003254, and FWA00005831. The total duration of the experiment was approximately one hour (Figure 1). Before the study, participants were informed about the study, and provided written informed consent. Participants were fitted with a BioHarness V.3 cardiovascular belt (Zephyr Technologies, USA) to record heart rate (HR), breathing rate (BR), and heart rate variability (HRV) to assess the activation of their autonomic nervous system during the disassembly and assembly task on MECCANO© reference 25207 (Figure 2). A five-minute physiological baseline measurement was conducted, during which participants remained at rest, seated for the entire five minutes. This baseline was necessary for measuring participants' HRV and comparing it with the results obtained in our specific conditions (Task Force of ESC&ASPE, 1996). Before the experiment began, participants familiarized themselves with the explanatory documentation for MECCANO© and performed a trial procedure not related to the experiment. This approach aimed to standardize the baseline knowledge level among all participants. The participants were divided into two independent experimental groups. Depending on the conditions, participants were informed of the steps to follow. The instructions differed according to the:

- High Mental Workload (HMW): Participants were given a 30-minute time limit to complete the task, with the remaining time visible. This threshold was determined from dry-run tests with three external participants who required approximately 20 minutes to complete the task; the 30-minute limit ensured feasibility. The HMW task included four additional pieces and three extra procedural steps (Appendix I).

- Low Mental Workload (LMW): Participants performed a less complex version of the task (four steps) without additional pieces and without a time constraint (Appendix I).

Given the effect of physiological activation depending on the time of day and the effect of time pressure on vigilance, the two groups of participants were evenly distributed between the morning and the afternoon (Roeser et al., 2012).

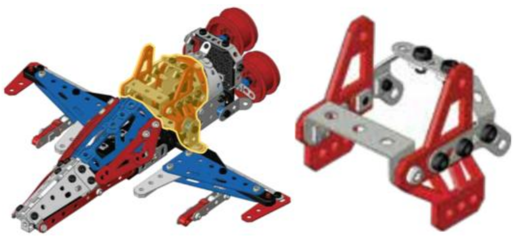


Figure 2. MECCANO© model – The yellow highlighted area in the left picture of the figure represents the part that was disassembled and assembled during the experiment.

For the HMW condition, the independent variables were designed to induce a high mental effort state, potentially influenced by factors such as time pressure and task complexity, aiming to replicate the demands of challenging cognitive situations (Santos and Melicio, 2019; Sheikhalishahi, Pintelon and Azadeh, 2016; Galy, 2017). Our objective was to induce an overall increase in mental workload rather than to isolate the independent effects of time pressure and task complexity (Galy, 2017). This is especially true in maintenance, where time pressure applied to complex tasks can lead to suboptimal mental states (Hobbs, 2021; Yiannakides and Sergiou, 2019). These deliberately imposed constraints primarily aim to identify which indicators were able to detect a high mental workload in a disassembly and assembly task. Furthermore, this study, conducted in a controlled laboratory setting, allowed us to better isolate and measure the effect of mental workload during our experiment. This will serve as a foundation for us to better understand and implement a more representative study within a real-world aeronautical maintenance environment. By keeping participants seated, physical workload was reduced so that the manipulation focused on mental workload. In addition, the acoustic environment was quiet, and the experiment was conducted in a room at a normal temperature. The participants followed their procedures by completing the disassembly steps, and once finished, they applied the assembly steps in reverse order (Appendix I). After task completion, participants completed the NASA-TLX self-assessment questionnaire (Hart and Staveland, 1988).

B. Measurement methods

B.1. Subjective measure

The NASA-TLX rating scale consists of six subjective subscales that assess various elements of the task on a scale from 0 to 100. This multidimensional scale is divided into six distinct parts, each with its own subscale: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (P), Effort (E), Frustration (F), and Overall Workload (OW). For the OW, it was determined using a pairwise comparison procedure. After assigning scores to the six dimensions, each participant was presented with pairs of dimensions (e.g., 'Mental Demand' vs. 'Physical Demand') and asked to select the one that contributed more to their overall workload (Hart and Staveland, 1988). In this study, participants used the NASA-TLX Multi-Equation Tool (Morales et al., 2020). This process was repeated until all dimensions had been compared with each other. Finally, the average of all weighted OW scores was calculated. Participants were guided by the experimenter to help them understand the different categories of workload.

B.2. Performance measure

The measured performance variable was the number of errors made during the experiment only on the steps performed by both groups. Errors such as forgetting disassembly and/or assembly, assembly in the wrong place, part inversion, and respecting the order of the procedure were measured in both conditions through video recording during the task.

B.3. Cardiovascular measure

As specified, for the cardiovascular data, one participant was excluded from the group due to the consumption of a psychoactive substance, and two participants' raw data were lost during the administration. So, for the cardiac data, the HMW group consisted of 17 participants, and the LMW group consisted of 16 out of 18 participants. HR, BR, and HRV were recorded during the experiment. All cardiovascular data analysis was performed using electrocardiogram (ECG) time data, but only the last five minutes of the HMW and LMW conditions were analyzed, as industry guidelines recommend measuring short-term HRV in this way (Task Force of ESC&ASPE, 1996; Castaldo et al., 2017; Tiwari et al., 2020). Furthermore, we chose to focus on the last five minutes of each condition to compare similar stages of the procedure. For the HMW condition, we chose to analyze this period as we estimated that time pressure would have the greatest impact on HRV indicators. Raw ECG signals were sampled at 1000 Hz; respiratory data were sampled at 1 Hz. Physiological data were exported in CSV format. Concerning HRV data, R–R intervals were corrected using the low-artifact correction tool in Kubios HRV Scientific Lite (version 4.1); artefactual beats were replaced using cubic-spline interpolation (Tarvainen et al., 2014). HRV analysis was performed in Kubios (Ognev et al., 2019). HRV measures in the frequency domain and time domain were calculated. Frequency domain indices included absolute high-frequency power (0.15Hz - 0.4 Hz; ms^2; HF), normalized high-frequency power (HFnu), absolute low-frequency power (0.07Hz - 0.14Hz; ms^2; LF), normalized low-frequency power (LFnu), and the ratio of LF and HF band powers (rLF/HF). Time domain measures included mean heart rate (MHR, beats per minute), mean R-R interval (MRR, ms), mean maximum heart rate value (Max HR, beats per minute), square root of the mean squared differences between adjacent normal R-R intervals (RMSSD, ms), standard deviation of R-R intervals (SDNN, ms), Baevsky's Stress Index (SI) reflects centralization of heart-rate control and is associated with sympathetic activity (Quendler, Trieb, and Nimmerichter, 2017; Ognev et al., 2019). Cardiovascular data were compared between HMW, LMW, and the baseline (BASELINE) recorded at experiment start. HRV assessment was performed according to generally accepted industry guidelines (Task Force of ESC&ASPE, 1996).

B.4. Normality and homoscedasticity tests

Statistics were conducted using R software (R Core Team, 2021). Regarding the NASA-TLX scores, for the HMW condition, the Shapiro-Wilk test revealed that the scores were not normally distributed for the OW sources ($p <.05$), except for the MD, PD, TD, P, E, and F sources ($p >.05$). For the LMW condition, the Shapiro-Wilk test revealed that the scores were not normally distributed for the MD, PD, TD, P and OW workload sources ($p <.05$), except for the E and F source ($p >.05$). To avoid Type I errors, which are a risk when the assumptions of a parametric test are not met, we performed a multivariate Permanova test. This non-parametric test was chosen because the Shapiro-Wilk test revealed that the NASA-TLX scores for most of the workload sources were not normally distributed. Multivariate permanova test allowed us to compare our groups without assuming normality while taking into account the joint variation of our NASA-TLX dimensions. As a post-hoc test, we used the Mann-Whitney test to compare the NASA-TLX workload sources because it's robust against non-normal data. However, for the F workload source, a t-test was performed because Levene's test showed no significant difference ($p > .05$).

For the performance data on the number of errors made, the Shapiro-Wilk test revealed that the data were not normally distributed for both the HMW and LMW conditions ($p < .05$). The Mann-Whitney tests were used to compare the number of errors between the two conditions.

We used the Shapiro-Wilk test to check for normality in the HRV data across the HMW, LMW, and BASELINE conditions. If the data for a condition was normally distributed ($p > .05$), we then performed a Levene's test on those conditions (Table 1). To avoid Type I errors, we performed a multivariate Permanova test to compare our groups while taking into account the joint variation of our HRV indicators. For a post-hoc analysis of indicators not following a normal distribution, we used a Mann-Whitney test to compare the cardiovascular indicators affected by mental workload. A t-test was performed on indicators where Levene's test showed no significant difference ($p > .05$).

III. Results

A. Demographic and sedentary results

For the demographic and sedentary data between the HMW and LMW groups, the Shapiro-Wilk test revealed that the scores were normally distributed for the weekly physical activity data ($p > .05$), except for Age, Alcohol consumption, Tobacco consumption, and psychoactive substance intake ($p < .05$). Therefore, for the weekly physical activity data, Levene's test showed no significant difference ($p > .05$), and the t-test also showed no significant difference ($t\ (-0.08) = 33.867$, $p = .934$). The Mann-Whitney test for independent samples revealed no significant difference for Age (*W*

=176.5, p =.652), Alcohol consumption (*W =153, p =.653*), tobacco consumption (*W =144, p =.163*), and psychoactive substance intake (*W =153, p =.345*). Therefore, our two groups show no significant differences in demographic and sedentary data.

B. Subjective results

The Permanova test revealed a significant score (*p<0.01*) between the set of NASA-TLX dimensions and conditions. As a post-hoc test, the Mann-Whitney test for independent samples used to compare the perceived workload sources between the conditions revealed differences but except for F (Figure 3). The effects of TD (*W =249, p =.005*) and E (*W =248.5, p =.006*) were highly significant between the conditions. The effects of MD (*W =230.5, p =.030*), PD (*W =242.5, p =.010*), and OW (*W =232.5, p =.026*) were significant between the conditions. However, the effects of P (*W =200.5, p =.227*) and F (*t (1.6262) =32.494, p =.113*) showed no significant difference between the conditions. As a result, the HMW group generally experienced a heavier workload than the LMW group.

Table 1. STATISTICAL PRE-TEST FOR HRV INDICATORS

	HMW	LMW	BASELINE	HMW vs LMW	HMW vs BASELINE	LMW vs BASELINE
HRV indicators	*Shapiro-Wilk (p-value)*			*Levene (p-value)*		
Mean RR (ms)	0.522	0.671	0.371	0.124	0.307	0.441
Mean HR (beats/min)	0.151	0.922	0.944	0.446	0.290	0.049 *
Max HR (beats/min)	0.745	0.856	0.157	0.454	0.420	0.090
SDNN (ms)	0.803	0.224	0.334	0.683	7.99E-04 ***	3.00E-03 ***
RMSSD (ms)	0.719	0.774	0.603	0.237	6.54E-05 ***	3.00E-03 ***
Stress Index	0.035 *	0.024 *	0.064	N/A	N/A	N/A
LF (ms²)	0.021	0.0001 ***	2.67E-05 ***	N/A	N/A	N/A
HF (ms²)	0.185	0.012 *	0.003 **	N/A	N/A	N/A
LF (nu)	0.110	0.409	0.312	0.416	0.220	0.042 *
HF (nu)	0.072	0.842	0.325	0.267	0.428	0.052
Ratio LF/HF	0.001 ***	0.266	2.02E-07 ***	N/A	N/A	N/A

Normality of the data was tested for all HRV indicator conditions. Then, the Levene test assessed the variance for data following the normal distribution. The p-values are indicated by asterisks.

*p < .05 **
*p < .01 ***
*p < .001 ****
N/A: Non applicable

C. *Performance results*

The Mann-Whitney test for independent samples used to compare the number of errors made between the conditions revealed a significant difference (Figure 4). The HMW group generated significantly more errors (*M =2.88, SD =2.16*) than the LMW group (*M =1.55, SD =2.25; W =232, p =.0246*). The effect of high mental workload triggered a higher number of errors during the disassembly and assembly task.

D. *Cardiovascular results*

Table 2 presents the HRV indicators between the HMW, LMW, and BASELINE conditions, while Table 3 presents the statistical results (p value). The Permanova test revealed a significant score ($p < 0.001$) between the set of HRV indicators and conditions. As a post-hoc test, the t-test for independent samples used to compare the Max HR indicator revealed only a significant difference between HMW (*M =103.18, SD =8.7*) and LMW (*M =95.75, SD =10.5; t (29.161) =2.2076, p =.035*). For the rest of the indicators, no significant difference was observed between the HMW and LMW conditions. However, when comparing to the BASELINE, interesting trends can be identified among the p-values for HMW vs. BASELINE and LMW vs. BASELINE, this is especially evident when comparing the HMW and LMW conditions for the MRR and MHR indicators, where the difference between HMW and LMW for MRR is *50 ms*, and the difference for MHR is *4.76 beats per minute*. Overall, the effect of high mental workload did not trigger a significant difference among all HRV indicators.

The Mann-Whitney test for independent samples used to compare the respiratory rate between the HMW (*M = 18.32, SD = 3.54*) and LMW (*M =16.89, SD =2.99; W =162, p =.3632*) conditions did not reveal a significant difference (Figure 5). Similarly, when compared to the BASELINE, the t-test did not show a difference between the p-values for LMW (*M = 16.89, SD = 2.99*) with BASELINE (*M =13.11, SD =3.66; t (36.273) =-5.0539, p =1.252E-05*) and for HMW (*M =18.32, SD =3.54*) with

BASELINE (*M =13.11, SD =3.66; W =69, p =1.049E-06*). The effect of high mental workload did not trigger a significant difference in respiratory rate.

All descriptive data for the complete set of results are available in Appendix II.

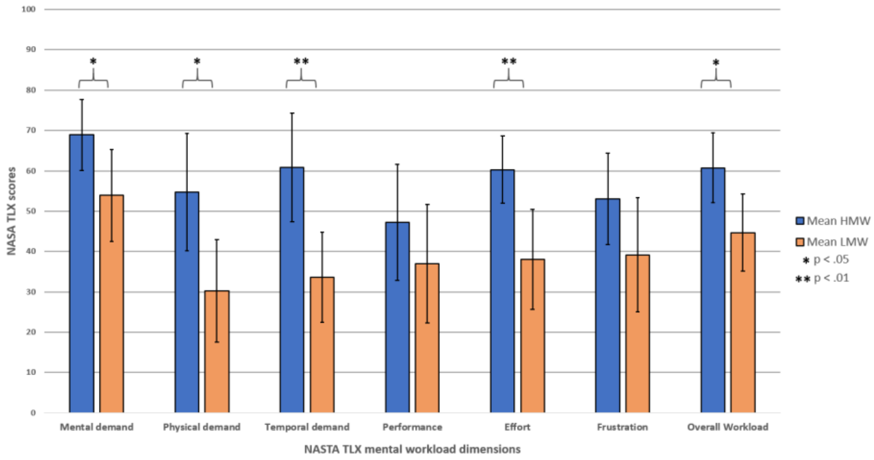


Figure 3. Histogram of the subjective workload measure conducted using NASA-TLX.

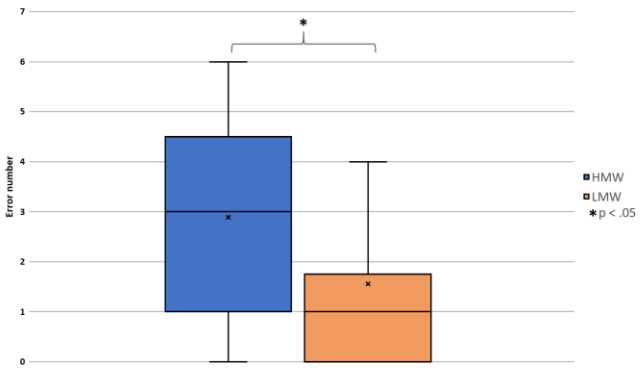


Figure 4. Boxplot of the number of errors made by participants across groups. The crosses represent the mean and the lines represent the medians.

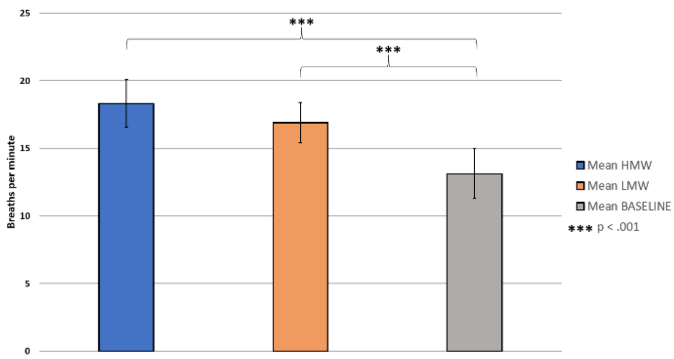


Figure 5. Histogram of the breathing rate across conditions.

Table 2. CARDIOVASCULAR INDICATORS DURING THE EXPERIMENT

Here are the HRV indicators during the last 5 minutes of the task for the HMW and LMW conditions. The data are presented as mean (SD). In the time domain, mean heart rate

HRV indicators	HMW (Mean ± SD)	LMW (Mean ± SD)	BASELINE (Mean ± SD)
MRR (ms)	697.5 ± 63.7	743 ± 94.8	839.49 ± 81
MHR (beats/min)	86.76 ± 8.4	82 ± 10.2	72.13 ± 6.8
Max HR (beats/min)	103.18 ± 8.7	95.75 ± 10.5	90.4 ± 7.4
SDNN (ms)	34 ± 9.4	35.025 ± 10.5	65.03 ± 22
RMSSD (ms)	25.94 ± 7.6	27.26 ± 10.3	55.7 ± 22.1
Stress Index	14.07 ± 4	13.55 ± 4.4	7.68 ± 2.4
LF (ms²)	561.82 ± 378.9	551.5 ± 482.2	2779.3 ± 2648
HF (ms²)	417.88 ± 306.7	366 ± 267.4	1438.7 ± 1094.6
LF (nu)	47.85 ± 13	47.64 ± 10.5	61.77 ± 17.35
HF (nu)	33 ± 14.4	31.55 ± 10.8	37.2 ± 17.4
Ratio LF/HF	2.45 ± 2.1	1.78 ± 0.9	2.72 ± 3

(MHR), mean R-R interval (MRR), mean maximum heart rate (Max HR), root mean square of successive differences (RMSSD), standard deviation of NN intervals (SDNN), and Baevsky's stress index (SI) were measured. In the frequency domain, absolute high frequency power (HF), absolute low frequency power (LF), normalized high frequency power (HFnu), normalized low frequency power (LFnu), and the ratio of LF to HF power (rLF/HF) were assessed

Table 3. STATISTICAL ANALYSIS RESULTS OF HEART RATE VARIABILITY INDICATORS

HRV indicators	HMW vs LMW p-value	HMW vs BASELINE p-value	LMW vs BASELINE p-value
Mean RR (ms)	0.119	3.69E-08 ***	0.0017 **
Mean HR (beats/min)	0.155	1.19E-06 ***	0.0012 **
Max HR (beats/min)	0.035 *	1.62E-05 ***	0.080
SDNN (ms)	0.761	1.73E-06 ***	4.33E-06 ***
RMSSD (ms)	0.683	6.50E-06 ***	1.91E-05 ***
Stress Index	0.732	4.45E-07 ***	2.86E-06 ***
LF (ms²)	0.787	5.45E-06 ***	2.29E-07 ***
HF (ms²)	0.691	3.70E-05 ***	3.90E-05 ***
LF (nu)	0.96	0.0027 **	0.0069 **
HF (nu)	0.745	0.367	0.169
Ratio LF/HF	0.957	0.823	0.619

All HRV indicators were tested using Mann-Whitney tests and t-tests. The p-values are indicated by asterisks.
*$p < .05$ **
*$p < .01$ ***
*$p < .001$ ****

IV. Discussion

The present study analyzed the effect of mental workload on a disassembly and assembly task involving cognitive resources similar to those of a maintenance activity. Our hypotheses were based on the ability of our indicators to detect a high mental workload on a disassembly and assembly task ~~similar to maintenance~~ in a laboratory context. Consequently, we expected an increase in the number of errors directly related to high mental workload. We also anticipated a pattern of subjective workload indicating significant differences between different sources of workload. Finally, we expected significant changes in cardiovascular data, which are indices of autonomic nervous system activity.

A. NASA-TLX and mental workload

The majority of the workload sources in the NASA-TLX showed significant differences between the two conditions. This highlights the impact of external constraints on the workload perceived by individuals during our experimentation. The task difficulty and time pressure resulted in significantly higher scores for time demand and effort for the HMW group compared to the LMW condition. They had to expend much greater cognitive resources in order to respond to the temporal objective and task complexity of the condition. Similarly, the mental and physical demands were significantly different for the HMW group. These results confirm the link between physical and mental workload during a disassembly and assembly task (Sugiharto, 2019; Dehais et al., 2020; Barajas-Bustillo et al., 2022; Morales et al., 2023). Time pressure and a challenging task, in addition to causing significant mental workload, can also intensify the physical workload felt by the operator. Understanding and diagnosing this link would represent a further advance in the design of safer maintenance systems. Regarding frustration, no statistically significant difference was identified. This could be attributed to similar frustration experienced by both groups due to the manipulation of small components and restrictive tools. This also contributes to the significant score of physical demand between the groups because the task difficulty and time pressure accentuated the constraining aspect of handling. Furthermore, although the NASA-TLX performance score did not show statistical significance, it nevertheless indicates a decline of performance for the HMW group, as indicated by the higher number of errors committed. The NASA-TLX tool once again confirms its excellent diagnostic ability regarding workload sources (Cegarra and Chevalier, 2008; Longo et al., 2022). In our case, it precisely identified the constraints faced by the operators and the sources of workload during our study through an effective measurement of overall mental workload.

B. Human error and mental workload

Associating performance issues with activities can reveal weaknesses in a system (Peach and Visser, 2020). Human error in many complex systems often indicates a design that is not human-centered, rendering these systems more vulnerable from a safety standpoint. Decreased performance can also be attributed to high mental workload during a task, stemming from factors such as an individual's limited capacities or the neurological mechanisms involved in regulating their goals (Bernard et al., 2021). This phenomenon can be readily observed in our study, as there was a clear deterioration in performance during the disassembly and assembly task. Errors such as incorrect installations, misplaced installations, and component inversions were much more prevalent in the HMW condition. In addition to the higher number of errors in the HMW condition, we can also observe a link with the Performance dimension of the NASA-TLX, where the score for the HMW condition is higher. This indicates lower performance satisfaction compared to the LMW condition, confirming our second hypothesis. Thus, our study represents a further step in identifying and predicting the most likely human errors in contexts demanding high mental effort. Therefore, establishing indicators of human performance that enhance individual performance and contribute to enhancing safety in the future is essential (Hobbs and Williamson, 2002; Bernard et al., 2020; Paxion, Galy and Berthelon, 2014).

C. *Cardiovascular data and mental workload*

Our objective was to understand the cardiovascular impacts of high mental workload, particularly examining the responses of HRV and BR indicators. HRV measurements, which can be correlated with mental workload, provide information about the respective actions of the SNS and PNS, leading to physiological changes in individuals (Young et al., 2014; Kostenko, 2017; Kramer, 2020). The temporal and frequency domain reveals the predominance of the PNS or SNS that may result from mental, emotional, and/or stress-related activity (Li et al., 2022; Delliaux et al., 2019). Among all the HRV indicators investigated in this study, only the Max HR temporal indicator identified a significant difference between the two groups, indicating that the HMW group generally reached higher HR levels during the task. However, in the temporal domain, we can observe clear trends between the groups for the indicators MRR, MHR, SDNN, and RMSSD. Overall, we noted a significant decrease in HRV indicators in both temporal and frequency domains compared to the BASELINE, without a clear prevalence of PNS or SNS components of the autonomic nervous system between the two groups. This indicates that these measures are sensitive to activity, although not sufficiently to detect differences in levels of difficulty in this context. Although the literature generally accepts that an increase in mental workload would stimulate the SNS and/or inhibit the PNS (Kostenko, 2017; Delliaux et al., 2019; Fan et al., 2020), the limited number of significant differences in HRV indicators in our study could be explained by a disassembly and assembly task that does not highlight a difference between two similar tasks where only the number of tasks and time pressure are varied. Additionally, according to Delliaux (2019), there might have been a phenomenon of anxious anticipation in the LMW group due to the laboratory conditions, potentially stimulating the SNS through HRV measurements for the HMW group and hindering the distinction between the two groups. Similarly, for BR, no significant disparity was found between the two groups, although clear trends in the number of breaths per minute were observed. Therefore, based on these results, our hypotheses have not been confirmed within the confines of our laboratory environment for the cardiovascular data. In conclusion, all these results confirm the sensitivity of cardiovascular measurements in detecting high mental workload. However, they also confirm the low diagnostic capacity of cardiovascular measurement tools (Cegarra and Chevalier, 2008; Li et al., 2022).

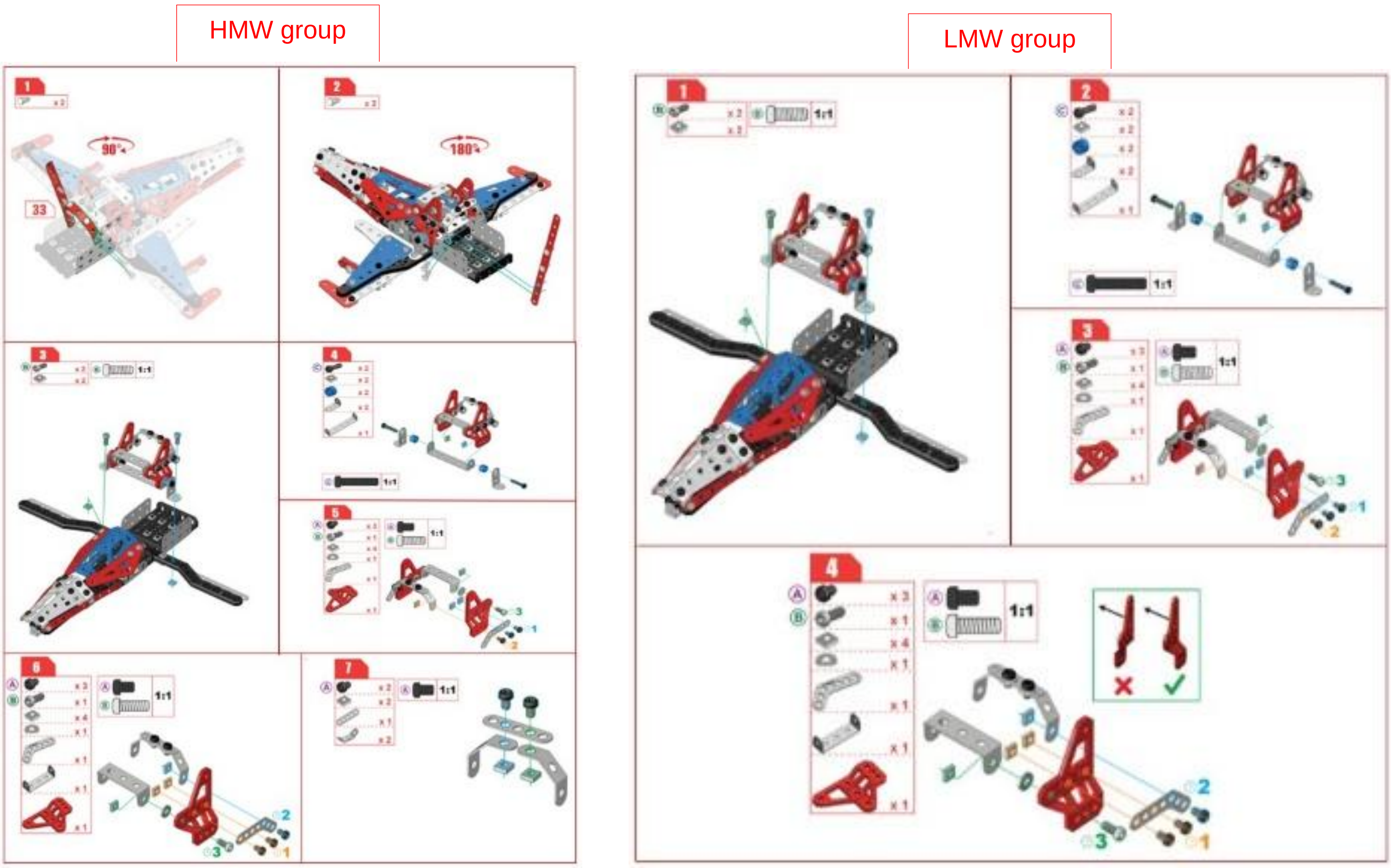
V. Conclusion

Our results suggest that this combination of multimodal measurements of mental workload may be effective in measuring high mental workload during a laboratory-based disassembly and assembly task that simulates maintenance. Although participants were subjected to representative maintenance constraints—such as following a procedure, time pressure, and tool manipulation—the study would benefit from more ecological validity by using expert maintenance participants and an aircraft maintenance task. In our study, only two categories of mental workload measures—subjective and performance—were able to identify and diagnose the constraints leading to a deterioration in performance. Otherwise, our cardiovascular indicators did not identify differences despite visible trends. Each tool individually has the potential to clarify and diagnose the experienced mental workload, but when combined, these data can help us enhance our understanding of human cognitive behavior. The knowledge developed from this study will aim, in a subsequent phase, to test this combination of mental workload measurement in a real maintenance environment. This future study will allow us to go further in experimenting with this combination of mental workload measures in a real maintenance context, where procedures can be more complex, the use of specific tools is required, and the consequences for the safety of both the aircraft and the operator can be fatal. Therefore, it is crucial for maintenance system designers to become more aware of the limitations of human fallibility and the unpredictable nature of many maintenance tasks (Peach & Visser, 2020). In this perspective, integrating HFE into maintainability and developing this multimodal method will be a first step towards reducing human errors and preserving system safety.

VI. Limits and perspectives

This study has some limitations. As it was conducted in a laboratory setting, it is necessary to verify if this method is applicable in a real maintenance context. Additionally, not all participants were experts and discovered the task to be performed on the day of the experiment, which could have affected the collected cardiovascular measures. However, it can also impact the performance level, which is also closely tied to individual ability levels. Moreover, other physiological measures could have been interesting for assessing mental workload. Neurophysiological measures such as fNIRS or EEG could have been used in this study to obtain direct data from the central nervous system.

Appendix



Appendix I. Construction steps for the two conditions: the HMW group follows the steps on the left, and the LMW group follows those on the right. The different disassembly and assembly steps are indicated from 1 to 7 for the HMW group and from 1 to 4 for the LMW group.

References

Bustillos, M. A. B., Maldonado-Macías, A. A., García-Alcaraz, J. L., Arellano, J. L. H., & Sosa, L. A. (2022). Considerations of the Mental Workload in Socio-Technical Systems in the Manufacturing Industry: A Literature Review. Research Anthology on Changing Dynamics of Diversity and Safety in the Workforce, 66-84. DOI: 10.4018/978-1-6684-2405-6.ch005

Barajas-Bustillos, M. A., Maldonado-Macías, A. A., Serrano-Rosa, M. A., Hernandez-Arellano, J. L., Llamas-Alonso, L., & Balderrama-Armendariz, O. (2023). Impact of experience on the sensitivity, acceptability, and intrusive of two subjective mental workload techniques: The NASA TLX and workload profile. Work, 75(4), 1265-1275. https://doi.org/10.3233/WOR-21132

Bernard, F., M. Zare, J. C. Sagot, and R. Paquin. 2020. "Using Digital and Physical Simulation to Focus on Human Factors and Ergonomics in Aviation Maintainability." Human Factors 62 (1): 37–54. https://doi.org/10.1177/0018720819844454.

Bernard, F., M. Zare, C. Murie, and J. C. Sagot. 2021. "Dimensions physiques et cognitives: vers une nécessaire prise en compte en maintenabilité aéronautique." Archives des Maladies Professionnelles et de l'Environnement 82 (2): 170–183. https://doi.org/10.1016/j.admp.2020.10.004.

Berthon, L., F. Bernard, S. Fleury, R. Paquin, and S. Richir. 2024. "Mental Workload: A Prerequisite for Future Maintenance Design." In Human Factors and Simulation. AHFE (2024) International Conference. AHFE Open Access, Vol. 139, edited by J. Wright and D. Barber. USA: AHFE International. https://doi.org/10.54941/ahfe1005025.

Cain, B. 2007. "A Review of the Mental Workload Literature." DTIC Document, July.

Castaldo, R., L. Montesinos, T. S. Wan, A. Serban, S. Massaro, and L. Pecchia. 2017. "Heart Rate Variability Analysis and Performance During a Repeated Mental Workload Task." In *EMBEC & NBC 2017: Joint Conference of the European Medical and Biological Engineering Conference (EMBEC) and the Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC), Tampere, Finland, June, 69–72. Springer Singapore. https://doi.org/10.1007/978-981-10-5122-7_18.

Cegarra, J., and A. Chevalier. 2008. "The Use of Tholos Software for Combining Measures of Mental Workload: Toward Theoretical and Methodological

Improvements.” Behav. Res. Methods 40: 988–1000. https://doi.org/10.3758/BRM.40.4.988.

Charles, R. L., and J. Nixon. 2019. “Measuring Mental Workload Using Physiological Measures: A Systematic Review.” Appl. Ergon. 74: 221–232. https://doi.org/10.1016/j.apergo.2018.08.028.

Dehais, F., A. Lafont, R. Roy, and S. Fairclough. 2020. “A Neuroergonomics Approach to Mental Workload, Engagement and Human Performance.” Front. Neurosci. 14. https://doi.org/10.3389/fnins.2020.00268.

Delliaux, S., A. Delaforge, J.-C. Deharo, and G. Chaumet. 2019. “Mental Workload Alters Heart Rate Variability, Lowering Non-linear Dynamics.” Front. Physiol. 10. https://doi.org/10.3389/fphys.2019.00565.

Dhillon, B. S. 1999. “Engineering Maintainability: How to Design for Reliability and Easy Maintenance.” In Engineering Maintainability, 1st ed., 180–197. Houston, TX, USA: Gulf Professional Publishing.

Dhillon, B. S. 2009. “Human Reliability, Error, and Human Factors in Engineering Maintenance: With Reference to Aviation and Power Generation.” In Human Reliability, Error, and Human Factors in Engineering Maintenance, 1st ed. Boca Raton, FL, USA: CRC Press. https://doi.org/10.1201/9781439803844.

Fan, X., C. Zhao, X. Zhang, H. Luo, and W. Zhang. 2020. “Issue Title: The 8th International Conference on Biomedical Engineering and Biotechnology.” Technology and Health Care 28 (S1): 67–80. https://doi.org/10.3233/THC-209008.

Force, T. 1996. “Standards of Measurement, Physiological Interpretation and Clinical Use.” Circulation 93 (5): 1043–1065. https://doi.org/10.1161/01.CIR.93.5.1043.

Galy, E. 2017. “Consideration of Several Mental Workload Categories: Perspectives for Elaboration of New Ergonomic Recommendations Concerning Shiftwork.” Theor. Issues Ergon. Sci. 19 (4): 483–497. https://doi.org/10.1080/1463922X.2017.1381777.

Geovanini, G. R., Vasques, E. R., de Oliveira Alvim, R., Mill, J. G., Andreão, R. V., Vasques, B. K., ... & Krieger, J. E. (2020). Age and sex differences in heart rate variability and vagal specific patterns–Baependi heart study. Global heart, 15(1), 71. doi: 10.5334/gh.873

Grusenmeyer, C. 2023. “Maintenance: des activités à risques. Fiche pratique de sécurité.” INRS ED 123., FRANCE.

Hart, S. G., and L. E. Staveland. 1988. “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research.” In Advances in Psychology, Vol. 52, edited by P. A. Hancock and N. Meshkati, 139–183. North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9.

Hobbs, A. 2021. “Aircraft Maintenance and Inspection.” In International Encyclopedia of Transportation, Vol. 2, edited by R. Vickerman, 25–32. Elsevier.

Hobbs, A., and A. Williamson. 2002. “Skills, Rules and Knowledge in Aircraft Maintenance: Errors in Context.” Ergonomics 45 (4): 290–308. https://doi.org/10.1080/00140130110116100.

Kostenko, A. S. 2017. “Évaluation multidimensionnelle et dynamique de la maitrise de la situation par l’opérateur: création d’un indicateur temps réel de charge mentale pour l’activité de supervision de drones.” Doctoral dissertation, Université de Bretagne Sud, France.

Kramer, A. F. 2020. “Physiological Metrics of Mental Workload: A Review of Recent Progress.” Multiple Task Performance, 279–328.

Le Gonidec, N. 2022. “Conceptualiser et évaluer la charge mentale de salariés dans un contexte d'usage d'outils numériques : Le cas d'une entreprise de télécommunications.” Ph.D. dissertation, Université Côte d'Azur, France.

Leplat, J., and J. C. Sperandio. 1967. “Measurement of Work Load by the Technic of an Added Task.” L'annee Psychologique 67 (1): 255–277. https://doi.org/10.3406/psy.1967.27563.

Li, W., R. Li, X. Xie, and Y. Chang. 2022. “Evaluating Mental Workload During Multitasking in Simulated Flight.” Brain Behav. 12 (4): e2489. https://doi.org/10.1002/brb3.2489.

Longo, L. (2015). A defeasible reasoning framework for human mental workload representation and assessment. Behaviour & Information Technology, 34(8), 758-786. https://doi.org/10.1080/0144929X.2015.1015166

Longo, L., C. D. Wickens, G. Hancock, and P. A. Hancock. 2022. “Human Mental Workload: A Survey and a Novel Inclusive Definition.” Front. Psychol. 13: 883321. https://doi.org/10.3389/fpsyg.2022.883321.

Mallat, C., J. Cegarra, C. Calmettes, and R. L. Capa. 2020. “A Curvilinear Effect of Mental Workload on Mental Effort and Behavioral Adaptability: An Approach with the Pre-Ejection Period.” Human Factors 62 (6): 928–939. https://doi.org/10.1177/0018720819855919.

Mandrick, K., G. Derosiere, G. Dray, D. Coulon, J.-P. Micallef, and S. Perrey. 2013. "Utilizing Slope Method as an Alternative Data Analysis for Functional Near-Infrared Spectroscopy-Derived Cerebral Hemodynamic Responses." Int. J. Ind. Ergonom. 43 (4): 335–341. https://doi.org/10.1016/j.ergon.2013.05.003.

Martin, C., S. Hourlier, and J. Cegarra. 2013. "La charge mentale de travail : un concept qui reste indispensable, l'exemple de l'aéronautique." Le Travail Humain 76 (4): 285–308. https://doi.org/10.3917/th.764.0285.

Morales, Á. F. C., Arellano, J. L. H., Muñoz, E. L. G., & Macías, A. A. M. (2020). Development of the NASA-TLX multi equation tool to assess workload. International Journal of Combinatorial Optimization Problems and Informatics, 11(1), 50.

Campoya Morales, Á. F., Hernández Arellano, J. L., Maldonado Macías, A. A., Rodríguez, Y., & Ochoa-Ortiz, C. A. (2023). Comparing workload and performance during demanding push force and mental tasks. Work, 75(1), 303-314. https://doi.org/10.3233/WOR-220067

Nkosi, M., K. Gupta, and M. Mashinini. 2020. "Causes and Impact of Human Error in Maintenance of Mechanical Systems." In MATEC Web Conf., 05001.

Ognev, A. S., V. A. Zernov, E. V. Likhacheva, L. P. Nikolaeva, M. Y. Rudenko, D. D. Dymarchuk, D. S. Yesenin, P. A. Maslennikova, and N. V. Mizin. 2019. "Cardiometric Detection of Effects and Patterns of Emotional Responses by a Human Individual to Verbal, Audial and Visual Stimuli." Cardiometry (1): 79–86. https://doi.org/10.12710/cardiometry.2019.14.7986.

Paxion, J., E. Galy, and C. Berthelon. 2014. "Mental Workload and Driving." Front. Psychol. 5. https://doi.org/10.3389/fpsyg.2014.01344.

Peach, R. H., and J. K. Visser. 2020. "Measuring Human Factors in Maintenance: A Literature Review." South African Journal of Industrial Engineering 31 (4): 104–114. https://doi.org/10.7166/31-4-2325.

Quendler, E., K. Trieb, and A. Nimmerichter. 2017. "Validation of Automated Detection of Physical and Mental Stress During Work in a Hühnermobil 225." Ann. Agric. Environ. Med. 24 (2): 329–331. https://doi.org/10.5604/12321966.1235183.

R Core Team. 2021. R: A Language and Environment for Statistical Computing [computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Reason, J. 2020. “Maintenance-Related Errors: The Biggest Threat to Aviation Safety After Gravity?” In Aviation Safety, Human Factors-System Engineering-Flight Operations-Economics-Strategies-Management, 465–470. CRC Press. https://doi.org/10.3233/THC-209008.

Roeser, K., F. Obergfell, A. Meule, C. Vögele, A. A. Schlarb, and A. Kübler. 2012. “Of Larks and Hearts — Morningness/Eveningness, Heart Rate Variability and Cardiovascular Stress Response at Different Times of Day.” Physiol. Behav. 106 (2): 151–157. https://doi.org/10.1016/j.physbeh.2012.01.023.

Santos, L., and R. Melicio. 2019. “Stress, Pressure and Fatigue on Aircraft Maintenance Personnel.” International Review of Aerospace Engineering 12 (1): 35–45. https://doi.org/10.15866/irease.v12i1.14860.

Sheikhalishahi, M., L. Pintelon, and A. Azadeh. 2016. “Human Factors in Maintenance: A Review.” Journal of Quality in Maintenance Engineering 22 (3): 218–237. https://doi.org/10.1108/JQME-12-2015-0064.

Sugiharto, F. M. 2019. “The Relationship Between Mental Workload and Occupational Stress Among Aircraft Maintenance Officers at PT X Hubungan Beban Kerja Mental dengan Stres Kerja pada Tenaga Kerja Bagian Perawatan Pesawat di PT X.” The Indonesian Journal of Occupational Safety and Health 8 (2): 233–239. https://doi.org/10.20473/ijosh.v8i2.14860.

Tao, D., H. Tan, H. Wang, X. Zhang, X. Qu, and T. Zhang. 2019. “A Systematic Review of Physiological Measures of Mental Workload.” Int. J. Environ. Res. Public Health 16: 2716. https://doi.org/10.3390/ijerph16152716.

Tarvainen, M. P., J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-aho, and P. A. Karjalainen. 2014. “Kubios HRV – Heart Rate Variability Analysis Software.” Comput. Methods Programs Biomed. 113 (1): 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024.

Tiwari, A., R. Cassani, J.-F. Gagnon, D. Lafond, S. Tremblay, and T. H. Falk. 2020. “Prediction of Stress and Mental Workload During Police Academy Training Using Ultra-Short-Term Heart Rate Variability and Breathing Analysis.” In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 4530–4533. https://doi.org/10.1109/EMBC44109.2020.9175414.

Wickens, C. D. 2017. “Mental Workload: Assessment, Prediction and Consequences.” In Human Mental Workload: Models and Applications. H-WORKLOAD Jun.

2017., Communications in Computer and Information Science, Vol. 726, edited by L. Longo and M. Leva. Cham: Springer. https://doi.org/10.1007/978-3-319-61061-0_2.

Xie, B., and G. Salvendy. 2000. "Prediction of Mental Workload in Single and Multiple Tasks Environments." Int. J. Cogn. Ergon. 4 (3): 213–242. https://doi.org/1207/S15327566IJCE0403_3.

Yiannakides, D., and C. Sergiou. 2019. "Human Factors in Aircraft Maintenance." In Aircraft Maintenance, 1st ed., 88–120. CRC Press.

Young, M. S., K. A. Brookhuis, C. D. Wickens, and P. A. Hancock. 2014. "State of Science: Mental Workload in Ergonomics." Ergonomics 58 (1): 1–17. https://doi.org/10.1080/00140139.2014.956151.

Zaki, R., A. Barabadi, A. N. Qarahasanlou, and A. H. S. Garmabaki. 2019. "A Mixture Frailty Model for Maintainability Analysis of Mechanical Components: A Case Study." Int. J. Syst. Assur. Eng. Manage. 10: 1646–1653. https://doi.org/10.20473/ijosh.v8i2.14860.