ENHANCING WORKLOAD ASSESSMENTS FOR VALIDATION ACTIVITIES ASSOCIATED WITH DBA AND BDBA SCENARIOS

Aaron Derouin and Alice Salway
Canadian Nuclear Safety Commission
410 Laurier Avenue West,
Ottawa, Ontario K1R 5E3
Canada
aaron.derouin@canada.ca; alice.salway@canada.ca

ABSTRACT

After the Fukushima Daiichi accident, nuclear regulators around the world have required that power reactor licensees develop more extensive emergency mitigating responses and severe accident management provisions beyond the defense-in-depth measures for design basis accidents (DBAs), previously in place. Workload assessments represent common validation techniques that are used to demonstrate that workers are able to perform tasks without unacceptable performance degradation. High workload is known to induced stress and fatigue and may severely diminish a worker's capacity to perceive, recognize, and respond appropriately during emergency or unanticipated events, which may result in undesirable consequences. In estimating workload during emergency and severe accident scenarios, power reactor licensees tend to rely on subjective measures of workload, such as NASA-TLX. Due to reported mismatches in the literature between subjective and physiologically-derived estimates of workload, it is prudent to see what more can be done to improve the current state of practice, in the context of emergency and severe accident conditions.

To improve confidence in workload estimates, it is advocated that the nuclear industry integrate physiological-based measures into current practices by making use of 'on-body' or 'wearable' physiological sensors. In this paper, an overview of three different types of empirical workload approaches is provided. The advantages of wearable physiological sensors are considered in the context of extreme environments and occupations, with tangible examples including heat stress and pupillometry. Suggestions for a consensus forum on workload are provided and a research plan directed at improving the current practice of workload estimation is offered for consideration.

Key Words: Workload, Validation, Subjective, Physiological, Wearable Sensors

1 INTRODUCTION

Following the Fukushima Daiichi accident, nuclear regulators around the world have required that power reactor licensees develop more extensive emergency mitigating responses and severe accident management provisions beyond the defense-in-depth measures for design basis accidents (DBAs), previously in place. In Canada, as part of the process of enhancing the response capabilities in a beyond design basis accident (BDBA) licensees have been tasked with confirming that operator actions are possible during the deployment of emergency mitigating equipment (EME), and the use of EME guides (EMEGs) and severe accident management guides (SAMGs) [1]–[3]. To demonstrate the feasibility of the proposed EMEGs and SAMGs, it was expected that licensees demonstrate, through validation activities, that human actions are achievable in the conditions that workers may be required to perform.

In Canada, for DBA scenarios, REGDOC-2.10.1, ‘Nuclear Emergency Preparedness and Response’, specifies that licensees are to validate the emergency response plan and procedures [4]. As part of managing the effectiveness, usability, technical accuracy and scope of the required severe accident management program, licensees are expected to complete verification and validation activities of human and organizational performance aspects to provide assurance that human actions are possible within the context of emergency operating procedures and SAMGs [5]. Accordingly, in the execution of SAMG actions, “field operator performance and human-machine interface issues under hazardous environments and conditions should be identified and considered” [5]. Furthermore, “improvement of accident management should be achieved through the consideration and incorporation of results from research in human performance, including decision-making” [5].

Recently CSA N290-12, ‘Human Factors in Design for Nuclear Power Plants’ [6] was published. This standard specifies the purpose of human factors in design validation for the ‘Human System Interface’ (HSI), as follows: to confirm that the HSI design supports the plant's safety and operational goals; to identify issues or problems which arose during the validation; to provide recommendations for design solutions; and to identify the impacts on procedures, staffing, and training. Though technology is designed to reduce workload, sometimes “the tasks required to operate the technology may actually increase workload which may, in turn, degrade human performance” [7].

To assess the cognitive and physical demands associated with the execution of emergency operating procedures, EMEGs, and/or SAMGs, licensees tend to rely on subjective measures of workload, such as NASA-TLX [8]. However, relying solely on subjective measures to assess workload may be problematic. For example, evidence from the literature suggests that an individual’s perception of workload and his/her relative physiological response do not always match. In trained subjects, from a physical exercise perspective, the perception of heat strain was found to be lower than a physiologically derived heat strain index obtained during exercise induced heat stress [9]. In addition to the decrease in physical work capacity, heat stress has been shown in the literature to have an effect on working memory and information processing. As well, the wearing of nuclear-biological-chemical protective clothing has been shown to significantly decrease tolerance time to exercise [10], and the same effect would be expected for similar protective clothing worn in the nuclear industry. Given that heat stress can be readily induced while wearing protective clothing, there is reason to consider the ability of nuclear workers to reliably perform time-sensitive field actions under adverse environmental conditions [9]. For these reasons, work schedules during the post-acute response at Fukushima were modified in consideration of heat stress [11].

High workload is known to induce stress as well as fatigue, which may severely diminish one’s capacity to perceive, recognize, and respond to emergency or unanticipated events and, thus, can place both the operator and system at risk [12]. Given the emergence of evidence [13]–[24] suggesting interaction effects between the different dimensions of workload (e.g. the impairment of physical performance from mental fatigue [13]), it is apparent that more could be done within the nuclear industry to understand the effects of these potential interactions on workload, under adverse and/ or extreme environmental conditions.

This paper explores the use of on-body physiological sensors, in conjunction with subjective measures for estimating the multidimensional aspects of workload in both emergency and severe accident scenarios in the nuclear industry. During field action responses that are required in emergency and accident scenarios, workers are required to interact with digital and analog human machine interfaces, in addition to performing physically demanding tasks. Therefore, a holistic view of workload is desirable. As well, a more comprehensive assessment of workload for control room operators during DBA and BDBA scenarios may be achieved using physiological-based measures. Additionally, a research plan aimed at unravelling the complexity of workload interaction effects under extreme conditions will be presented.

2 BACKGROUND ON WORKLOAD ASSESSMENT APPROACHES

The assessment of workload broadly falls into one of three categories: subjective measures, performance measures, and physiological measures.

2.1 Subjective Measures of Workload

Subjective measures rely on participants to quantify their experience of workload either during or immediately after completing a task [12]. Typically, subjective measures make use of questionnaires or measurement scales in attempts to assess multiple dimensions of workload, such as mental demand, physical demand, temporal demand, performance, and frustration.

2.2 Performance-Based Workload Measures

Performance-based measures of workload rely on examining the capacity of an individual during the execution of primary or secondary tasks. By measuring how well a person performs on the task with increasing workload, an estimate of mental workload can be determined [25]. In the literature, three types [26] of performance-based workload assessment have been identified, including: speed and accuracy; task analysis techniques; and the use of secondary tasks to examine aspects of primary task performance. Secondary task performance-based measures may be useful in laboratory settings, but their practicality and utility in assessing workload during field-based validation activities is questionable.

2.3 Physiologically-Based Workload Measures

Physiologically-based workload measures are able to record the continuous physical responses of the body, and thus are able to infer the amount of mental work [25] and physical work that a person is experiencing. In physiological-based measurement approaches, “subjects are monitored by an array of physiological sensors, some requiring contact with the subject’s body through electrolyte sensors (e.g., electroencephalography, EEG; electrocardiography, ECG) while others are standoff sensors (e.g., eye tracking device embedded in the physical system)” [12].

3 THE UTILITY OF WEARABLE PHYSIOLOGICAL SENSORS FOR VALIDATION ACTIVITIES

To provide sufficiently reasonable estimates of workload, it is crucial that more than one workload assessment approach is utilized throughout the various stages of validation. This advocated approach is in agreement with that of Tran et al. (2007) [12], who proposed that “physiological assessment should be used jointly with traditional behavioral measures (e.g., workload questionnaire)”.

3.1 Challenges with Overreliance on Subjective Measures of Workload

Human factors engineers/ specialists tend to rely on subjective measures to assess workload, and typically overlook physiologically-based measures. Invariably, “among the numerous mental workload measures, subjective assessment has received the most attention” [27], which is likely due to its ease of use. However, in the literature some authors have identified theoretical and practical suspicions with regard to subjective assessment [27]. According to Zhang and Yu (2010), a large number of studies have found many critical flaws using behavioural indices or subjective measures [28]. Some of the limitations associated with the sole reliance on subjective workload measures, include: an inability to provide moment-to-moment fluctuations of mental fatigue [28]; an inability “to capture ‘real-time’ workload change without provoking a task-interruption nuisance onto the subject” [12]; and suspicions related to the “inherent biases (e.g., social desirability) that comes with self-evaluation” [12] and “subjective vulnerability” due to confirmation biases [12]. According to Liu and Wickens (1987), subjective measures may be limited, based on the researchers understanding of workload, as follows:

> *Thus the reliable and sensitive measurement of subjective workload assessment is heavily determined by the construction and implementation of the specific rating scale being used. The construction of a workload rating scale is determined by the researcher’s understanding and definition of the concept of workload, and psychometric considerations* [27].

Although the disadvantages of subjective workload measures have been described in detail in the literature, more needs to be done to dispel the notion that, “subjective perceptions of cognitive effort may constitute the essence of workload” [27]. Regardless of how many dimensions of subjective workload there are or perceived to be, the assessment of workload in the absence of physiological data is prone to bias and overreliance on the memory [29], [30], of the worker rating the task.

3.2 Recent Advances in Physiological-Based Measures for Workload Assessment

Previous obstacles to using physiologically-based measures, such as its ‘obtrusive nature’ [12], can no longer be used as a reason to not include these more objective-based measures. An emerging area of sports science [31] and human performance [32]–[34] includes the use of wearable physiological sensors to estimate workload across various workload dimensions. The advent of wearable physiological sensors or on body sensors has virtually eliminated previously perceived limitations thought to restrict their use in the assessment of workload. Accordingly, “physiological measures have become an increasingly popular approach in assessing workload of newly developed system design” [12]. The reason for this increasing popularity may be due to the fact that “physiological measurements are not constrained to several limitations that are inherent in self-report questionnaires and performance modeling” [12]. This category of workload assessment is able to provide an exact and objective measure of the body’s physical reaction to workload and stress, and does not require a direct verbal response from the person. Physiological measures permit a more objective workload assessment and can provide “real-time” evaluation, thus allowing the system designer to quickly and accurately identify task- and equipment-related problems as they occur. Furthermore, analysts are able to time-stamp the physiological assessment results with that of operator’s behavioral performance, which allows for finer precision in identifying high risk occurrences during task performance [12]. As well, physiological-based measures foster the assessment of ‘emergent’ behaviors, thus allowing designers to detect emergent and unanticipated design issues [12]. Furthermore, different physiological indicators have been shown to reflect different aspects of task demand [35]. However, physiological-based measures do require some proficiency in the application and interpretation [35] of the most suitable and robust analytic techniques required to identify valid and tangible changes to an individual's workload state. To this end, several hardware and software packages have been developed by different groups to facilitate the analysis of human performance and workload [36].

Recent evidence from the literature suggests that wearable physiological sensors may be able to detect the early onset of heat stress [37]–[40]. This is an important consideration in the context of accident and emergency scenarios, as workers that are called on to respond may face adverse environmental conditions, including high heat and humidity during the execution of physically demanding work activities. Beyond the interaction effects of heat stress and cognitive function, more recent research is beginning to reveal additive interaction effects between other dimensions of workload, such as time pressure and task difficulty [41].

In DBA and BDBA scenarios, the management of high levels of arousal is vital to the successful performance of tasks under mental stress. In extreme operational environments, Costanzo and Hatfield (2013) [42] have suggested that state sensitive biomarkers (i.e., EEG, heart rate variability, etc.) could be used to classify if a human operator is an adaptive state, in terms of neural or psychomotor efficiency. By examining elite performers in sports, such as intercollegiate athletes or Olympians, a better "understanding of the neural basis for such abilities to adaptively cope with stressful events" [42] may be gleaned. "More specifically, elite athletes may be uniquely resilient to stress perturbation through the ability to regulate emotions" [42]. The ability to manage or regulate emotion [42] is an essential determinant of the quality of performance, under the pressure of extreme conditions which is highly dependent on the individual's appraisal or perception of the stimulus rather than of the objective stimulus itself [42]. Emerging evidence from both the neuroscience and imaging disciplines is providing great insight into the respective use of pupillometry and thermal imaging [43] for the assessment of cognitive workload, in adverse conditions. Although, the use of pupillometry in the assessment of workload [44]–[47] is not new, "the study of changes in pupil diameter has attracted considerable attention" [48]. Small changes in the diameter of the pupil are linked to cognitive processes associated with the activation of an area of the brainstem called the locus coeruleus [49]. This area of the brain has been linked with the release of norepinephrine and pupil dynamics. Recent studies have demonstrated that changes in pupil size are positively related to learning [50], [51] and to challenges in task performance [52].

In various domains including nuclear, several researchers have observed pupil dilation in response to increases in workload [53], [54]. Recent work by Bhavsar et al. (2015) [53] found, in 44 participants, that pupillometry captured distinct changes in pupil diameter that were linked to abnormal process situations, in which the control room operators were required to respond.

Collectively, this line of research suggest that the measurement of pupil dynamics may provide great insight into optimal performance and the adaptive state of the operator during recognition, decision making and response processes required in emergency and severe accident scenarios. This approach may prove to be useful in characterizing individual variability with respect to optimal performance in extreme environments, and in quantifying the degree to which individuals are specifically suited to performing optimally in demanding environments [55]. Conversely, the use of pupillometry could be used to capture design issues or address gaps in training.

Though pupillometry and other eye gaze tracking metrics are able to measure visual attention and provide an indication of workload, most eye trackers have a limited tracking range (e.g. ± 35 degrees in the horizontal direction) [56]. In wide field of view (FOV) scenarios, the pupils may no longer be visible in the eye trackers' view scope [56]. Therefore, an integration of head pose and eye gaze data is suggested in the context of wide FOV human machine interfaces used in primary and secondary control rooms.

In the context of emergency and accident exercise scenarios, it is important to gather as much of the 'right type' of objective workload information as possible to ensure that workers are adequately supported and not overloaded from a workload perspective. Assuring that workload estimates are reasonable strengthens the licensees' response capabilities and reliability in the execution of work activities required during events that may range in severity, from anticipated operational occurrences up to and including severe accidents.

4 THE IMPORTANCE OF REASONABLE WORKLOAD ESTIMATES IN EXTREME ENVIRONMENTS

From a human reliability perspective, a more refined deterministic understanding of human performance under extreme conditions is desirable to realistically estimate workload associated with response capabilities in severe accident scenarios. When modelling performance effectiveness in conditions of extremely high stress, considerable uncertainty relates to the human error probability (HEP), if stress threat conditions persist beyond 30 minutes [57]. If personnel have been trained to mitigate the effects of the accident, it is assumed that their performance reliabilities start to recover, but for untrained personnel the HEP remains at around 0.25 [57]. Additionally, the uncertainty bounds in persistent threat stress level conditions around the 0.25 HEP estimate are wide-ranging, from 0.05 to 1.0 [57], which means that some workers perform well and others make errors every time. Task performance reliability is also dependent on the nature of the tasks, i.e., step-by-step versus dynamic [57]. To add further complexity and uncertainty, the interaction effects of physiological and cognitive stressors on workload under extreme conditions are largely unknown.

In a broader view of human performance in extreme conditions, experts from different disciplines and industries met at the 2014 Nuclear Energy Agency's Working Group on Human and Organizational Factors and identified several knowledge gaps, including the potential need for simulation of field or local tasks in harsh environments [58]. Assurance of the response capabilities of power reactor licensees in different accident scenarios could be enhanced by collecting and analyzing objective workload data across various workload dimensions during large scale exercises and integrated with the growing knowledge of human performance under extreme conditions.

In the literature, an emerging need has been identified "to utilize both qualitative and quantitative information to catalog the performance of humans in safety critical and extreme environments" [59]. Barnett and Kring (2003) define an extreme environment as "settings that possess extraordinary technological, social, and physical components that require significant human adaptation for successful interaction and performance" [59]. A framework for the extreme environment taxonomy was established based on a comprehensive review of literature conducted by Barnett and Kring, and includes nuclear power plant operations, which was classified as an extreme occupation. In four of the listed extreme environments and occupations (space, aviation, polar and surgery/emergency room), subject matter experts (SMEs) were asked to rate the degree and frequency of 28 factors within three main categories of extremes: physical, physiological, and psychological. Though nuclear was not rated by SMEs, preliminary results from this work suggest that major similarities exist between the specific environments. However, given that the interaction effects of physiological and cognitive stressors on workload under extreme conditions are largely unknown, it is prudent that the nuclear industry, in the wake of the Fukushima Daichi accident, takes practical steps to better understand how these interaction effects influence workload and human reliability, in the context of both accident and emergency scenarios. To this end, research on human performance within the nuclear context should make use of integrated and networked physiological sensors in conjunction with subjective workload measures to inform decisions regarding workload in power reactor operations under extreme emergency conditions. In the last decade, numerous integrated physiological sensor suits/ systems have come on the market that offer the ability to simultaneously record physical, physiological, and cognitive workload data. However, many of these systems have yet to be validated against more robust empirically-based laboratory measures and techniques.

5 CONSIDERATIONS FOR THE ADOPTION OF WEARABLE PHYSIOLOGICAL SENSORS

Part of the challenge in measuring workload lies in the fact that "workload means different things to different people — for some workload is thought of "as something physical while others believe workload to be more about mental activity or time pressure" [26]. From an examination of the literature it is evident that a dichotomy exists amongst human factors practitioners and their individual perceptions on what constitutes, defines, and characterizes workload. (i.e. mental workload vs. physical workload). For example, Shriram et al. (2013), explains that, "physiological measurement relies on evidence that increased mental demands lead to increased physical response from the body" [25], yet provides no indication of the utility of physiologically-based measures for understanding physical workload stress and/or physiological strain. To overcome this potential barrier a consensus must be reached to establish an appropriate definition that considers all categories and dimensions of workload. For example, in September 2000, a Consensus Forum [60] in relation to establishing Bona Fide Requirements for Physically Demanding Occupations was held in Canada. A similar consensus forum could be planned and held to achieve a common and shared understanding of the next steps in establishing a common definition of workload and the development of a research plan.

In establishing a research plan, a collaborative multi-disciplinary approach is required to harmonize research efforts in the areas of human performance and workload assessment under a common goal. Collaborative efforts between human factors, neuroscience, biomechanics, engineering, exercise physiology, and the like, will only prove to be fruitful if organized in a pragmatic manner. To this end, work should be directed at addressing the following:

- Characterization of power reactor accidents as an extreme environment using the 3 categories of extremes (physical, physiological, and psychological) and the 28 factors identified by Barnett and Kring [59]:
- Development of an overview of sensor types and their respective utility for evaluating extremes within physical, physiological and psychological workload domains;
- Validation of sensor data against different physiological measures, and against subjective and/or performance-based measures;

- Integration of data and real time monitoring with subjective measures;
- Analysis of the availability, feasibility, and utility of off-the-shelf integrated wearable sensor arrays for use within the nuclear operation settings;
- Interference concerns with sensitive equipment; and
- An overview of ethical issues and concerns associated with the use of wearable physiological sensors

6 CONCLUSIONS

Over the last several decades, a great deal has been learned, from a scientific perspective, about how humans react to specific physical and psychological demands [26]. However, according to Casner and Gore [26], "the measurement and evaluation of workload is far from the exact science we would like it to be". To address this issue head on, increased recognition is needed with respect to the inherent challenges and limitations associated with measuring that which is perceived and in quantifying the capacity of some of the more intangible dimensions of workload, from recorded physiological data. These challenges are best summarized, as follows:

> *"There is no workload measurement scale or technique that offers the same reliability as the scales used to measure height, weight, pressure or other physical quantities. Similarly our efforts to rigorously quantify the point at which human operators reach their workload 'boiling point' fall far short of those that quantify similarly important states of physical matter* [26].

In the context of extreme environments this perceived limitation of workload evaluation needs to be overcome. The inclusion of on-body or wearable physiological sensors during emergency and severe accident validation activities, drills and exercises, alongside the currently used subjective measures, can increase the defensibility and understanding of workload assessment in the nuclear industry.

7 REFERENCES

[1] Canadian Nuclear Safety Commission, "*CNSC Integrated Action Plan On the Lessons Learned From the Fukushima Daiichi Nuclear Accident*," Ottawa, (2013).

[2] Canadian Nuclear Safety Commission, "*INFO-0828, CNSC Staff Action Plan on the CNSC Fukushima Task Force Recommendations*," Ottawa, (2012).

[3] Canadian Nuclear Safety Commission, "*CNSC Staff Action Plan on the CNSC Fukushima Task Force Report and CNSC Management Response*," Ottawa, (2012).

[4] Canadian Nuclear Safety Commission, *REGDOC-2.10.1, Nuclear Emergency Preparedness and Response*. PWGSC catalogue number CC172-115/2014E-PDF ISBN 978-1-100-25029-8, Ottawa, (2014).

[5] Canadian Nuclear Safety Commission, *REGDOC-2.3.2 v2 - Accident Management*. PWGSC catalogue number CC172-116/2015E-PDF ISBN 978-0-660-03338-9, (2015).

[6] Canadian Standards Association Group, *N290.12-14, Human factors in design for nuclear power plants*. Mississauga, (2014).

[7] T. B. Aldrich, S. M. Szabo, and C. R. Bierbaum, "The Development and Application of Models to Predict Operator Workload During System Design," in *Applications of Human Performance Models to System Design*, Boston, MA: Springer US, pp. 65–80, (1989).

[8] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Advances in Psychology*, **vol. 52**, 1988, pp. 139–183 pp. 139–183, (1988).

[9] P. Tikuisis, T. McLellan, and G. Selkirk, "Perceptual versus physiological heat strain during exercise-heat stress," *Med. Sci. Sport. Exerc.*, **vol. 34**, no. 9, pp. 1454–1461, (2002).

[10] T. M. McLellan and Y. Aoyagi, "Heat strain in protective clothing following hot-wet or hot-dry heat acclimation.," *Can. J. Appl. Physiol.*, **vol. 21**, no. 2, pp. 90–108, Apr. (1996).

[11] K. Wada, T. Yoshikawa, T. Hayashi, and Y. Aizawa, "Emergency response technical work at Fukushima Dai-ichi nuclear power plant: occupational health challenges posed by the nuclear disaster.," *Occup. Environ. Med.*, **vol. 69**, no. 8, pp. 599–602, Aug. (2012).

[12] T. Q. Tran, R. L. Boring, D. D. Dudenhoeffer, B. P. Hallbert, M. D. Keller, and T. M. Anderson, "Advantages and disadvantages of physiological assessment for next generation control room design," in *2007 IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting*, pp. 259–263, (2007).

[13] S. M. Marcora, W. Staiano, and V. Manning, "Mental fatigue impairs physical performance in humans," *J. Appl. Physiol.*, **vol. 106**, no. 3, pp. 857–864, Mar. (2009).

[14] A. DiDomenico and M. A. Nussbaum, "Effects of different physical workload parameters on mental workload and performance," *Int. J. Ind. Ergon.*, **vol. 41**, no. 3, pp. 255–260, May (2011).

[15] A. DiDomenico and M. A. Nussbaum, "Interactive effects of physical and mental workload on subjective workload assessment," *Int. J. Ind. Ergon.*, **vol. 38**, no. 11–12, pp. 977–983, Nov. (2008).

[16] R. D. Moore, M. W. Romine, P. J. O'connor, and P. D. Tomporowski, "The influence of exercise-induced fatigue on cognitive function," *J. Sports Sci.*, **vol. 30**, no. 9, pp. 841–850, May (2012).

[17] T. McMorris, J. Sproule, A. Turner, and B. J. Hale, "Acute, intermediate intensity exercise, and speed and accuracy in working memory tasks: A meta-analytical comparison of effects," *Physiol. Behav.*, **vol. 102**, no. 3–4, pp. 421–428, Mar. (2011).

[18] V. Labelle, L. Bosquet, S. Mekary, T. T. M. Vu, M. Smilovitch, and L. Bherer, "Fitness Level Moderates Executive Control Disruption during Exercise Regardless of Age," *J. Sport Exerc. Psychol.*, **vol. 36**, no. 3, pp. 258–270, Jun. (2014).

[19] K. Lambourne and P. Tomporowski, "The effect of exercise-induced arousal on cognitive task performance: A meta-regression analysis," *Brain Res.*, **vol. 1341**, pp. 12–24, Jun. (2010).

[20] V. Labelle, L. Bosquet, S. Mekary, and L. Bherer, "Decline in executive control during acute bouts of exercise as a function of exercise intensity and fitness level," *Brain Cogn.*, **vol. 81**, no. 1, pp. 10–17, Feb. (2013).

[21] Y. K. Chang, J. D. Labban, J. I. Gapin, and J. L. Etnier, "The effects of acute exercise on cognitive performance: A meta-analysis," *Brain Res.*, **vol. 1453**, pp. 87–101, May (2012).

[22] C. Perry, M. Sheik-Nainar, N. Segall, R. Ma, and D. Kaber, "Effects of physical workload on cognitive task performance and situation awareness," *Theor. Issues Ergon. Sci.*, **vol. 9**, no. 2, pp. 95–113, Mar. (2008).

[23] A. Mazloumi, F. Golbabaei, S. Mahmood Khani, Z. Kazemi, M. Hosseini, M. Abbasinia, and S. Farhang Dehghan, "Evaluating Effects of Heat Stress on Cognitive Function among Workers in a Hot Industry," *Heal. Promot. Perspect.*, **vol. 4**, no. 2, pp. 240–6, (2014).

[24] A. Dietrich and P. B. Sparling, "Endurance exercise selectively impairs prefrontal-dependent cognition," *Brain Cogn.*, **vol. 55**, no. 3, pp. 516–524, Aug. (2004).

[25] R. Shriram, M. Sundhararajan, and N. Daimiwal, "EEG Based Cognitive Workload Assessment for Maximum Efficiency," *IOSR J. Electron. Commun. Eng.*, **vol. 7**, pp. 34–38, (2013).

[26] S. M. Casner and B. F. Gore, "*Measuring and Evaluating Workload: A Primer*," NASA/TM—2010-216395, (2010).

[27] Y. Liu and C. D. Wickens, "*Mental workload and cognitive task automation: an evaluation of subjective and time estimation metrics*," Engineering-Psychology Research Laboratory, University of Illinois at Urbanna-Champaign, prepared for NASA Ames Research Center, Moffett Field, California, Contract No. NASA-NAG-2-38, (1987).

[28] C. Zhang and X. Yu, "Estimating mental fatigue based on electroencephalogram and heart rate variability," *Polish J. Med. Phys. Eng.*, **vol. 16**, no. 2, pp. 67–84, Jan. (2010).

[29] L. Hope, W. Lewinski, J. Dixon, D. Blocksidge, and F. Gabbert, "Witnesses in Action: The Effect of Physical Exertion on Recall and Recognition," *Psychol. Sci.*, **vol. 23**, no. 4, pp. 386–390, Apr. (2012).

[30] W. F. Moroney, D. W. Biers, and F. T. Eggemeier, "Some Measurement and Methodological Considerations in the Application of Subjective Workload Measurement Techniques," *Int. J. Aviat. Psychol.*, **vol. 5**, no. 1, pp. 87–106, Jan. (1995).

[31] D. A. James and N. Petrone, *Sensors and Wearable Technologies in Sport*. Springer Singapore,

Singapore, (2016).

[32] R. Parker, A. Vitalis, R. Walker, D. Riley, and H. G. Pearce, “Measuring wildland fire fighter performance with wearable technology,” *Appl. Ergon.*, **vol. 59**, pp. 34–44, Mar. (2017).

[33] J. A. Healey and R. W. Picard, “Detecting Stress During Real-World Driving Tasks Using Physiological Sensors,” *IEEE Trans. Intell. Transp. Syst.*, **vol. 6**, no. 2, pp. 156–166, Jun. (2005).

[34] F. Seoane, I. Mohino-Herranz, J. Ferreira, L. Alvarez, R. Buendia, D. Ayllón, C. Llerena, and R. Gil-Pita, “Wearable Biomedical Measurement Systems for Assessment of Mental Stress of Combatants in Real Time,” *Sensors*, **vol. 14**, no. 4, pp. 7120–7141, Apr. (2014).

[35] U.S. Nuclear Regulatory Commission, *Integrated System Validation: Methodology & Review Criteria. NUREG/CR-6393,* (1998).

[36] A. F. Rabbi, K. Ivanca, A. V. Putnam, A. Musa, C. B. Thaden, and R. Fazel-Rezai, “Human performance evaluation based on EEG signal analysis: A prospective review,” in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1879–1882, (2009).

[37] M. Baratchi, L. Teunissen, P. Ebben, W. Teeuw, J. Laarhuis, and M. van Steen, “Towards decisive garments for heat stress risk detection,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp ’16*, pp. 1095–1100, (2016).

[38] C. A. Holm, L. Pahler, M. S. Thiese, and R. Handy, “Evaluation of physiological strain in hot work areas using thermal imagery,” *J. Therm. Biol.*, **vol. 61**, pp. 8–15, Oct. (2016).

[39] P. Pancardo, F. D. Acosta, J. A. Hernández-Nolasco, M. A. Wister, and D. López-de-Ipiña, “Real-Time Personalized Monitoring to Estimate Occupational Heat Stress in Ambient Assisted Working.,” *Sensors*, **vol. 15**, no. 7, pp. 16956–80, Jul. (2015).

[40] E. Gaura, J. Kemp, and J. Brusey, “Leveraging Knowledge From Physiological Data: On-Body Heat Stress Risk Prediction With Sensor Networks,” *IEEE Trans. Biomed. Circuits Syst.*, **vol. 7**, no. 6, pp. 861–870, Dec. (2013).

[41] E. Galy, M. Cariou, and C. Mélan, “What is the relationship between mental workload factors and cognitive load types?,” *Int. J. Psychophysiol.*, **vol. 83**, no. 3, pp. 269–275, Mar. (2012).

[42] M. E. Costanzo and B. D. Hatfield, “Brain Biomarkers of Neural Efficiency during Cognitive-Motor Performance: Performing under Pressure,” Springer, Berlin, Heidelberg, pp. 123–132, (2013).

[43] J. Stemberger, R. S. Allison, and T. Schnell, “Thermal Imaging as a Way to Classify Cognitive Workload,” in *2010 Canadian Conference on Computer and Robot Vision*, pp. 231–238, (2010).

[44] J. Beatty, “Pupillometric measurement of cognitive workload,” Technical report #11, Office of Naval Research, ADA050036, UCLA, (1977).

[45] M. Schwalm, A. Keinath1, and H. D. Zimmer, “Pupillometry as a method for measuring mental workload within a simulated driving task,” in *Human Factors for assistance and automation*, S. Publishing, Ed. Maastricht, the Netherlands, pp. 1–13, (2008).

[46] O. Palinko, A. L. Kun, A. Shyrokov, and P. Heeman, “Estimating cognitive load using remote eye tracking in a driving simulator,” in *Proceedings of the 2010 Symposium on Eye-Tracking Research*

& Applications - ETRA '10, p. 141, (2010).

[47] R. Lew, B. P. Dyre, T. Soule, S. A. Ragsdale, and S. Werner, "Assessing Mental Workload from Skin Conductance and Pupillometry using Wavelets and Genetic Programming," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, **vol. 54**, no. 3, pp. 254–258, Sep. (2010).

[48] D. Rozado and A. Dunser, "Combining EEG with Pupillometry to Improve Cognitive Workload Detection," *Computer (Long. Beach. Calif).*, **vol. 48**, no. 10, pp. 18–25, Oct. (2015).

[49] B. Laeng, S. Sirois, and G. Gredebäck, "Pupillometry," *Perspect. Psychol. Sci.*, **vol. 7**, no. 1, pp. 18–27, Jan. (2012).

[50] M. R. Nassar, K. M. Rumsey, R. C. Wilson, K. Parikh, B. Heasly, and J. I. Gold, "Rational regulation of learning dynamics by pupil-linked arousal systems," *Nat. Neurosci.*, **vol. 15**, no. 7, pp. 1040–1046, Jun. (2012).

[51] R. Cohen Hoffing and A. R. Seitz, "Pupillometry as a Glimpse into the Neurochemical Basis of Human Memory Encoding," *J. Cogn. Neurosci.*, **vol. 27**, no. 4, pp. 765–774, Apr. (2015).

[52] P. R. Murphy, I. H. Robertson, J. H. Balsters, and R. G. O'connell, "Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans," *Psychophysiology*, **vol. 48**, no. 11, pp. 1532–1543, Nov. (2011).

[53] P. Bhavsar, B. Srinivasan, and R. Srinivasan, "Pupillometry Based Real-Time Monitoring of Operator's Cognitive Workload To Prevent Human Error during Abnormal Situations," *Ind. Eng. Chem. Res.*, **vol. 55**, no. 12, pp. 3372–3382, Mar. (2016).

[54] Q. Gao, Y. Wang, F. Song, Z. Li, and X. Dong, "Mental workload measurement for emergency operating procedures in digital nuclear power plants," *Ergonomics*, **vol. 56**, no. 7, pp. 1070–1085, Jul. (2013).

[55] M. P. Paulus, E. G. Potterat, M. K. Taylor, K. F. Van Orden, J. Bauman, N. Momen, G. A. Padilla, and J. L. Swain, "A neuroscience approach to optimizing brain resources for human performance in extreme environments," *Neurosci. Biobehav. Rev.*, **vol. 33**, no. 7, pp. 1080–1088, Jul. (2009).

[56] H. Cai and Y. Lin, "An integrated head pose and eye gaze tracking approach to non-intrusive visual attention measurement for wide FOV simulators," *Virtual Real.*, **vol. 16**, no. 1, pp. 25–32, Mar. (2012).

[57] F. Bercha, C. Brooks, and F. Leafloor, "Human performance in arctic offshore escape, evacuation and rescue," in *The Proceedings of the 13th International Offshore and Polar Engineering Conference,* Honolulu, Hawaii, USA, (2003).

[58] NEA/CSNI/R(2015)16, "Human performance under extreme environments with respect to a resilient organization," in *Proceedings of a Committee on the Safety of Nuclear Installations International Workshop*, (2015).

[59] J. S. Barnett and J. P. Kring, "Human Performance in Extreme Environments: A Preliminary Taxonomy of Shared Factors," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, **vol. 47**, no. 8, pp. 961–964, Oct. (2003).

[60] N. Gledhill, *Bona Fide Occupational Requirements - Proceedings of the Consensus Forum on Establishing BONA FIDE Occupational Requirements*, (2000).