

# Workload Is Multidimensional, Not Unitary: What Now?

Gerald Matthews<sup>(✉)</sup>, Lauren Reinerman-Jones, Ryan Wohleber,  
Jinchao Lin, Joe Mercado, and Julian Abich IV

Institute for Simulation and Training (IST), University of Central Florida (UCF),  
Orlando, FL, USA

{gmatthews, lreinerm, rwohlebe, jlin,  
jmercado, jabich}@ist.ucf.edu

**Abstract.** It is commonly assumed that workload is a unitary construct, but recent data suggest that there are multiple subjective and objective facets of workload that are only weakly intercorrelated. This article reviews the implications of treating workload as multivariate. Examples from several simulated task environments show that high subjective workload is compatible with a variety of patterns of multivariate psychophysiological response. Better understanding of the cognitive neuroscience of the different components of workload, including stress components, is required. At a practical level, neither subjective workload measures nor single physiological responses are adequate for evaluating task demands, building predictive models of human performance, and driving augmented cognition applications. Multivariate algorithms that accommodate the variability of cognitive and affective responses to demanding tasks are needed.

**Keywords:** Workload · Task demands · Psychophysiology · Electroencephalogram (EEG) · Electrocardiogram (ECG) · Stress · Performance · Individual differences

## 1 Introduction

Mental workload is most simply defined in terms of the allocation of processing resources to meet task demands [1]. It is acknowledged that both sources of demand and resources are multifaceted, but, as a recent review states, explaining workload “in terms of demand/resource balance offers an attractive and parsimonious approach to this otherwise multidimensional construct” (p. 2) [2]. Thinking of workload in relation to a single ratio, i.e., demands: resources, is also attractive to systems designers, as it allows the definition of a “redline” as the ratio approaches 1.0 [3]. Subjective workload assessments, such as the NASA-TLX [4], typically provide an overall workload score that may be used to inform systems evaluation and design.

But what if the unitary workload model is incorrect? Suppose that there is no scalar value that adequately captures workload - how can we then understand workload conceptually and apply workload assessments in practical ergonomics? In this chapter, we review some of the evidence that calls into question the utility of the unitary model,

taken primarily from recent studies conducted in our lab. We also consider the prospects for developing workload models in the absence of a unitary construct.

### 1.1 General Workload: An Elusive Construct

General unitary workload is an intuitively appealing concept, given that cognitive overload is a universal human experience. However, it may be questioned from a variety of perspectives. First, in experimental studies that vary task demands, different workload metrics may dissociate, calling into question the validity of one or more of the metrics concerned [5]. Usually, dissociations are interpreted as indicating measurement problems; for example, subjective measures may not be sensitive to task differences associated with data limitations. However, they may point towards more fundamental uncertainties over construct definition. Second, multiple resource theories have become increasingly popular as a means for explaining dual-task interference data [1, 6]. Such theories may be compatible with some general integration of the loads imposed on different pools of capacity [2], but questions remain about how to weight multiple load factors to derive overall workload [7]. Third, the conventional stress/strain metaphor for workload [2] assumes some passive deformation of cognitive functioning (strain) as the external load (stress) increases. However, there are various dynamic factors that the metaphor does not capture, such as people's attempts to actively manage workload [8] and regulate task goals [9], and the malleability of resources over various timespans [2, 10].

The most basic objection to the assumption of unity is lack of psychometric support. Human factors has a history of embracing broad-based constructs such as workload, stress and arousal because of their potential for synthesizing findings and design recommendations across a variety of task contexts. These approaches have often failed to fulfill their potential because of the lack of a psychometrically sound metric for the construct concerned. For example, research on general arousal effects on performance devolved into futile attempts to work backwards from stressor interactions to fit data to the Yerkes-Dodson Law inverted-U curve post hoc [11]. Workload measurement is on a stronger footing because there appears to be a general factor underpinning the subjective experience of workload. The six rating scales of the NASA-TLX typically intercorrelate, suggesting they measure a common construct. However, modern psychometric techniques for identifying latent factors such as confirmatory factor analysis and item response theory have been largely neglected in the workload domain.

Given the limitations of self-report, especially in high-stakes test settings, objective workload assessments are essential. A range of psychophysiological techniques have shown considerable promise (see [12] for a review). However, the literatures on leading techniques, such as those using the electrocardiogram (ECG), electroencephalogram (EEG), or recent brain hemodynamic methods, have tended to develop in isolation from one another. The few comparative tests that have been performed typically do not include sufficient participants for standard psychometric analyses, such as factor analysis, to be performed.

Psychometric studies of psychophysiological response are challenging for a range of reasons, including the expense of running substantial numbers of participants, and

the sensitivity of responses to extraneous physiological factors, such as general metabolic activity. Nevertheless, if workload is unitary, the assumption is that different metrics reflect some common latent factor. A basic difficulty is that standard workload criteria, such as sensitivity, diagnosticity, selectivity, and others [13] are designed (quite effectively) for evaluating outcomes of experimental studies in which task load manipulations produce changes in mean response across groups of participants. However, building a strong psychometric case requires analysis of inter-subject correlations of the various metrics to demonstrate a unitary latent factor, a standard neglected in conventional workload criteria [14].

## 1.2 Psychometrics of Physiological Workload Indices

Recently, we investigated the psychometric properties of multiple psychophysiological metrics secured from a fairly substantial sample of individuals ( $N = 150$ ) performing a simulation of unmanned ground vehicle (UGV) operation, under various task demands [14, 15]. We employed a suite of sensors selected to assess a range of previously validated workload responses [16]. Sensors included ECG and EEG, from which conventional workload metrics were extracted, such as heart-rate variability (HRV) and frontal theta. Hemodynamic indices of workload were obtained bilaterally from two sensors: functional near infrared spectroscopy (fNIR) which indexes prefrontal oxygen saturation, and transcranial Doppler sonography (TCD) which yields measures of cerebral blood flow velocity (CBFV) in the middle cerebral arteries. Eye tracking metrics including frequencies and durations of fixations were also obtained.

Twelve independent measures of each response were obtained from the different conditions, defined by task load factors such as task difficulty, multi-tasking needs, and event rate. These proved to be highly internally consistent. That is, we can measure a given response at the individual level reliably. As further discussed below, several responses were sensitive to task manipulations, such as single- vs. dual-task performance. However, while data generally met conventional workload criteria [13], responses from different sensor systems were mostly independent from one another, and there was no prospect for modeling a unitary latent factor. The lack of covariance of alternate measures represents a major failure in convergent validity. Furthermore, while the NASA-TLX was also sensitive to task demand manipulations, correlations between this subjective measure and psychophysiological response hardly exceeded chance levels resulting in strongly divergent subjective and objective metrics.

Standard psychometric models, such as those derived from confirmatory factor analysis and structural equation modeling, are made up of two parts [17]. First, there is a measurement model that attaches measured variables to latent factors. Second, there is a structural model that specifies the relationships between latent factors. The failure of convergence of alternate workload metrics might be attributed to failures in the measurement model. Response measures may be poor indicators of activity in the brain systems that control workload. There may also be variability in the measurement model across individuals, in line with the response specificity principle [18], e.g., HRV but not frontal theta might be sensitive to task demands in one person, and vice versa in

another. Studies identifying individualized workload classifier algorithms [19] support this possibility.

It is plausible that measurement model issues contribute to the lack of convergence, and improvements in psychophysiological recording and analysis may alleviate the problem in the future. However, there are also reasons to question whether workload response is controlled by a unitary neurocognitive system, and we may thus need to elaborate the structural model of workload components. Imposing external task loads on operators impose multiple, concurrent neurocognitive responses, including cortico-reticular arousal, changes in brain metabolic activity, specific processing routines, high-level executive processing, task- and state-directed effort and regulation of stress and coping [20]. While these various responses may commonly occur together, they may not be strongly correlated within the individual. That is, there is not a single ‘strain’ response. Instead, there are various layers of regulation of the response to task stimuli, including anticipation of task demands based on prior expectancies. Correspondingly, even relatively simple but demanding tasks such as vigilance are supported by a complex network of brain areas [21], which may operate somewhat independently in supporting different neurocognitive functions.

At this point, we emphasize that the ultimate value of treating workload as unitary remains open. Improved measurement technology and models, as well as accommodation of individual variability in response, may yet salvage the construct. The practical utility of general subjective workload assessments is also well-established [22]. However, we will proceed to explore the consequences of abandoning a unitary workload construct in favor of a multidimensional conceptualization, should this step prove to be necessary.

## 2 Multivariate Workload Evaluation

Extant criteria for evaluating workload metrics [13, 22] already contain an inbuilt tension between sensitivity and diagnosticity. The idea behind diagnosticity is that the workload measure should identify the source of workload, as the NASA-TLX does in distinguishing mental demands, physical demands and other sources. However, maximizing diagnosticity requires that the measure is insensitive with regard to task-irrelevant sources of workload, a requirement that may limit the overall sensitivity of the instrument. Existing measures may not adequately discriminate sources of workload directly tied to task demands (e.g., rate of stimulus input) from those that reflect workload management strategy (e.g., effort). For example, the Multiple Resources Questionnaire (MRQ) [6] is geared primarily towards demand factors that can be specified objectively, such as auditory and visual task load, whereas the NASA-TLX includes more elusive, psychologically-defined qualities such as effort and frustration.

A multivariate conceptualization of workload avoids such difficulties by treating each index as a distinct diagnostic metric. The NASA-TLX and other subjective metrics provide an estimate of the participant’s awareness of task demands. However, given that the NASA-TLX does a rather poor job of capturing physiological response [14], it does not provide a comprehensive ‘gold-standard’ workload assessment.

Contrasting findings from two recent studies [14, 23] illustrate the multivariate approach. One study ( $N = 81$ ) applied the sensor set previously described [16] (without eye tracking) to investigate workload response in simulated Generalized Pressurized Water Reactor Nuclear Power Plant (NPP) Main Control Room (MCR) operation. The participant, in the role of a Reactor Operator (RO), works with a Senior Reactor Operator (SRO) to perform tasks requiring (1) checking that controls are appropriately set, (2) detecting changes in gauges, and (3) implementing responses that changed the state of controls. Designing plants to provide manageable workloads requires attention to different task elements of these kinds.

NASA-TLX data suggested that workload was highest for detection, and lower for checking and response implementation. However, psychophysiological data suggested a more complex pattern. Response magnitudes were calculated as the percentage change from a no-task baseline. Figure 1 shows these data by task for selected measures. Several responses were insensitive to the task manipulation and are omitted from the figure. Generally, responses to workload were bilateral, and so most of the indices graphed are averaged across hemisphere: the exception was CBFV which showed a lateralized (left-hemisphere) response. EEG responses (spectral power densities: SPDs) were averaged across three frontal sites (F3, F4, Fz).

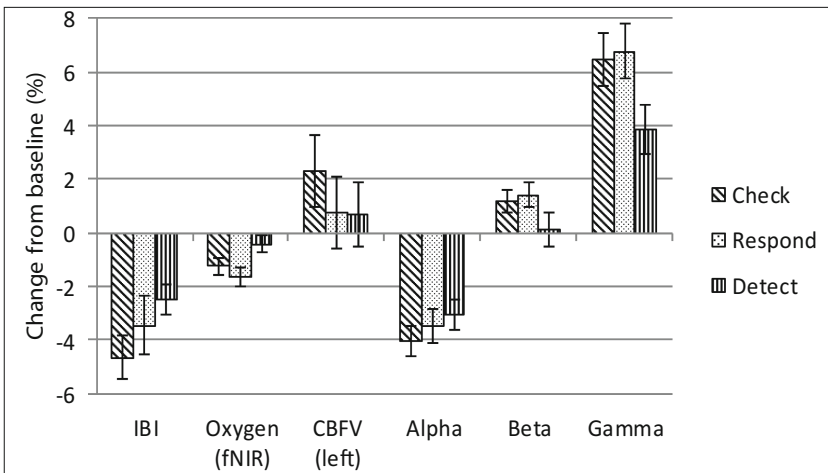


Fig. 1. Selected workload responses in three task conditions

Consistent with NASA-TLX data, bilateral prefrontal cortex blood oxygenation (fNIR) was higher for the detection task than the other two. By contrast, EEG data suggested that workload was lowest for detection. This task was associated with the highest level of alpha, but also with the lowest levels of high-frequency EEG (beta, gamma). Yet another pattern was implied by TCD and ECG data suggesting that workload was actually highest in the checking task. Left-hemisphere CBFV, measured by TCD, is elicited by demanding verbal tasks, including working memory [24, 25]. The checking task produced elevation of left-hemisphere CBFV, relative to the other

two tasks. Cardiac IBI was also lowest for this task, although HRV was insensitive to task differences. Thus, concordance between subjective and objective metrics was poor, and any conclusion about workload differences between the tasks depended entirely on the choice of metric. However, we can identify patterns of response that correspond to the unique demands of each task.

A second study [14, 15], introduced above, profiled workload response during simulated UGV operation. Two key manipulations were task type and multi-tasking. (A third manipulation, event rate, is beyond our current scope). In single-task conditions, the operator was required either to detect threatening human figures (enemy insurgents) in a video feed, or to detect changes in icons representing friendly and unfriendly forces on a map display. The subjective workload imposed by threat detection was substantially lower than that of change detection, perhaps because of the evolution of the visual system to process realistic human character stimuli. In multi-tasking conditions, participants performed both tasks simultaneously.

NASA-TLX data confirmed that both change detection and multi-tasking elicited substantially higher workload. Performance data tended to follow the TLX; for example, multi-tasking produced interference. However, the psychophysiological data again provided a more nuanced picture of workload response which contrasted with findings from the NPP study [23]. Several indices were diagnostic of task type but not multi-tasking. Specifically, change detection induced higher EEG theta (including frontal theta), lower HRV, higher prefrontal oxygenation (fNIR), and higher values for the pupillometric Index of Cognitive Activity (ICA). By contrast, multi-tasking did not affect any of these indices but it reduced eye fixation duration. As noted, TLX scores were not significantly correlated with any of the metrics that were sensitive to the manipulations, again calling into question the meaningfulness of any overall workload score. Speculatively, we might suggest that processing arbitrary symbols or icons (change detection) is more effortful than processing realistic human character stimuli. By contrast, with the UGV simulation, the need to exert cognitive control over scan patterns in order to monitor two task windows renders eye fixations the most sensitive metric for multi-tasking.

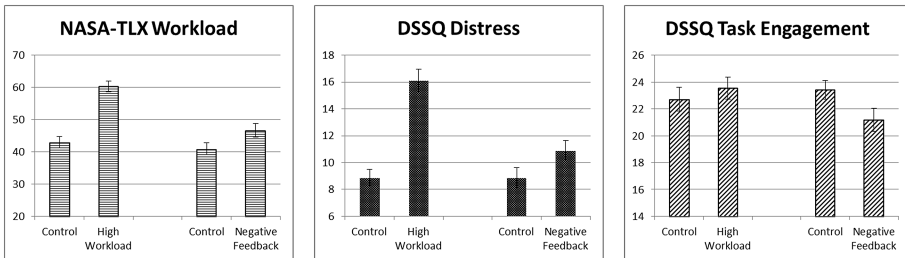
### 3 Extension to Task Stress

In a recent, unpublished study [26], we used the sensor suite [16] (less eye tracking) to examine responses to a simulated UAV task, in which task manipulations were expected to be stressful. UAV operation typically requires coordination of multiple subtasks, making the operator vulnerable to both overload and concerns about maintaining adequate performance [27]. One manipulation was intended to cause cognitive overload, by increasing the number of UAVs controlled from two to six, and reducing the time available to perform elements of the mission. A second manipulation involved the presentation of negative feedback messages suggesting that the person was failing to perform adequately (irrespective of actual performance).

The study also used the short Dundee Stress State Questionnaire (DSSQ) [28] to monitor subjective stress response. This multidimensional instrument assesses distress, task engagement, and worry. Typically, in performance studies, distress elicited by task

stressors follows subjective workload. For example, unpublished analyses for the simulated UGV study [14] showed that DSSQ distress was substantially higher both for change detection and for multi-tasking. In the UAV study, however, we considered that the qualitative difference between overload and negative evaluation as sources of stress might provoke differing multidimensional patterns of DSSQ response. The workload issue is then the extent to which both subjective stress response and psychophysiological response correspond to overall subjective workload as indexed by the NASA-TLX.

Subjective data from the study are shown in Fig. 2, confirming that the two tasks differ qualitatively in response pattern. These represent mean levels for each variable in paired low demand and high demand conditions. “Stress” associated with high workload took the form of highly elevated NASA-TLX scores and increased distress, with little change in task engagement. However, negative evaluation produced only modestly increased distress and workload, together with a decline in task engagement, implying that noncontingent negative feedback is demotivating.



**Fig. 2.** Effects of high workload and negative feedback on three subjective measures

Turning to the psychophysiological metrics, we found yet another pattern of response, differing from that found in both the NPP [23] and UGV [14] studies. The metrics that were sensitive to the task manipulations were high-frequency EEG power (beta and gamma), which increased, and HRV, which also increased. The two manipulations did not differ significantly in their impacts on these responses. The effects on high-frequency EEG suggest that simulated UAV operation may produce a rather subtle, “cognitive” stress response that is not expressed either in increased mental effort (e.g., frontal theta) or in classic arousal response (e.g., reduced alpha and IBI).

The increase in HRV was particularly unexpected, given that increased cognitive demand is traditionally considered to *lower* HRV [29]. A resolution to the paradox is suggested by a meta-analysis [30], which coordinated HRV response with activation of discrete brain areas in functional Magnetic Resonance Imaging (fMRI) studies. Higher HRV may signal increased top-down regulation of emotional processing in structures such as the amygdala. High frequency EEG might also reflect the processing associated with emotion-regulation. The impact of cognitive load on HRV may then depend on the extent to which task demands are perceived as threats requiring initiation

of emotion-regulation. However, at least in the present paradigm, the subjective data are necessary to differentiate the threat of cognitive overload from the threat of negative evaluation; in this respect, the DSSQ was more diagnostic than the psychophysiology. The notion of general workload appears inadequate to capture the nature of the similarities and differences of the response patterns elicited by the two task manipulations.

#### 4 Modeling Individual Differences in Performance

The standard demand/resource model [2] suggests that workload should be diagnostic of performance at the individual level, to the extent that the demand: resource ratio is indicative of poor performance. (Note that secondary task performance may be more sensitive than primary task performance to resource insufficiency [1]). However, existing research already shows that it is difficult to find any single index of resource utilization that is uniquely associated with performance. In relation to psychological measures, a previous study found that cognitive ability, task-focused coping, and task engagement all predicted vigilance task performance, but each had unique effects [20]. “Resources” appeared to be a construct that emerged from several personal characteristics, rather than being identified with any single measures. Similarly, a study showed that task engagement and phasic CBFV response were correlated but uniquely predictive of vigilance [24].

Unpublished data from the UGV simulation study [14, 15] suggest a similar conclusion. Performance levels varied across task conditions that differed in task demands. However, there were consistent individual differences in performance on the change detection and threat detection tasks. Thus, we can calculate average accuracy scores on these performance tasks, across the different conditions. Table 1 shows correlations between averaged workload and stress indices and these performance criteria. Omitted psychophysiological measures did not correlate with either performance index. Only change detection was related to psychophysiological as well as subjective metrics. Several indices of low workload (low TLX scores, three psychophysiological metrics) were linked to more accurate performance.

**Table 1.** Correlations between averaged performance accuracy and selected measures

Task	Subjective scales				Psychophysiology		
	NASA-TLX	Distress	Engagement	Worry	HRV	Beta	Fixation duration
Change detection	-.39**	-.39**	.45**	-.32**	.26**	-.30**	.26**
Threat detection	-.12	-.21**	.28**	-.33**	-.07	-.14	.16

Note. \*\* $p < .01$

A multiple regression showed that the seven predictor variables together explained 40.4 % of the variance in change detection (adjusted  $R^2 = .37$ ). With the four subjective



measures entered into the equation first, the three psychophysiological measures added significantly to the variance explained ( $\Delta R^2 = .08, p < .01$ ). Conversely, the subjective measures contributed significantly to prediction with the psychophysiological variables controlled ( $\Delta R^2 = .20, p < .01$ ). In the final equation, five predictors remained significant at  $p < .05$  or better (i.e., TLX, task engagement, HRV, fixation duration, beta). That is, we can identify individuals who are sufficiently vulnerable to overload that their change detection is impaired quite effectively, but we need multiple subjective and objective indicators to do so. No single index represents a “gold standard” for behavioral impairment resulting from overload. Indeed, the different measures may correspond to different psychological processes which work together to preserve or threaten performance competency [20].

## 5 Augmented Cognition

Greater validity in predicting individual performance may support augmented cognition applications, especially adaptive automation or intelligent tutors. The accurate assessment of workload is critical for creating closed-loop systems to augment performance [31]. Such systems require a valid workload metric to drive adaptive allocation of task functions between human and machine. In the absence of such a metric, adaptation of automation to match the cognitive demands being experienced is likely impossible. Similarly, in training applications, developing a system that effectively provides feedback relies on valid, near real-time workload assessment. Data such as those in Table 1 suggest that determining a valid metric for practical applications is not such a simple task as suggested by unitary resource theory, but instead might require individuation that is also task dependent and domain dependent. That is, implementing augmented cognition is not task and domain agnostic.

Acceptance of multiple workload components allows progress to be made. It remains true that psychophysiological metrics are more powerful than subjective ones for application. Within a given task domain, it is possible to develop multivariate algorithms for predicting performance vulnerability. For example, in the UGV operation scenario [14, 15], a person showing low HRV, high beta and short fixation durations may be especially vulnerable. This pattern of response might be used to drive an automated aid for the task concerned. Individuals may vary in relation to which indices are most diagnostic of compromised performance. One operator may be especially taxed by overload of executive processing, another by misdirection of effort, and yet another by stress. Multivariate modeling may allow adaptive automation to be directed towards the specific vulnerabilities of individual operators. Efforts at developing idiographic workload classifiers based on techniques such as neural net analysis show the promise of this approach, although defining cross-task classifiers remains problematic [19]. The limitation of such approaches is loss of diagnosticity; it may be difficult to determine *why* different workload processes are more or less predictive of performance across different individual operators.

## 6 Future Directions and Challenges

Table 2 summarizes the central dilemma for unitary workload assessment. Each row of the table corresponds to a task paradigm in which we have identified demand manipulations which influence NASA-TLX workload predictably. In addition, each manipulation is associated with poorer performance and higher workload (although it is difficult to compare performance levels across the NPP tasks [23]). However, at the physiological level, each workload response appears quite different, with each inducing a unique pattern of response. Indeed, in some cases, changes in metrics were unexpected, such as increased HRV in the high cognitive load condition of the UAV study [26], and reduced high frequency EEG during the NPP detection task. Increases in subjective workload are rather poorly diagnostic of changes in neural function.

Another challenge is that subjective and objective indices are not mutually reducible to one another. Both are necessary to optimize performance prediction [24]. In a research context, both may be validly assessed, but subjective assessments lose validity in high-stakes real-life settings where the person may be motivated to distort subjective response. It remains to be seen whether future developments in sensor technology will allow objective metrics to be effectively substituted for subjective ones.

Our principal argument is that, from a psychometric perspective, the structural model of workload cannot be identified with a single latent factor. Instead, there appear to be multiple latent dimensions, which should be identified with more narrowly defined characteristics than global workload. A challenge to better definition of multiple factors is uncertainty over the measurement model, i.e., the best response measures to identify each factor. It remains plausible that there is individual variation in the mappings of responses onto latent factors.

An applied approach that might address the limitations imposed by individual differences is to examine clustering patterns across individuals to determine if a subset of classificatory patterns emerge. Ideally, then, a person newly entering a system or task environment would complete a short baseline multivariate assessment of workload

**Table 2.** Task-sensitive workload metrics in four task paradigms

Study	Task or manipulation	Metric(s) for elevated workload (expected)	Metric(s) for reduced workload (unexpected)
NPP simulation [23]	Detection (vs. other tasks)	Frontal oxygenation (fNIR)	Lower beta Lower gamma Higher alpha
UGV simulation [14]	Change detection (vs. threat detection)	Higher frontal theta, lower HRV, prefrontal oxygenation (fNIR), higher ICA	
UGV simulation [14]	Dual-task (vs. single task)	Shorter fixation duration	
UAV simulation [26]	Cognitive demands (vs. low demands)	Higher beta Higher gamma	Higher HRV

response that enabled the correct pattern to be identified. Furthermore, it is not enough to assume that tasks that are similar theoretically or in processing structure produce an equivocal workload response, but rather the tasking environment (domain) must be considered when developing a closed-loop system. Accommodating the complexities of individual workload response may be critical for realizing the real-world potential of augmented cognition.

**Acknowledgements.** This work was in part supported by the Air Force Office of Scientific Research (AFOSR) (FA 9550-13-1-0016) and the US Army Research Laboratory (ARL) (W91CRB-08-D-0015). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of AFOSR, ARL or the US Government.

## References

1. Wickens, C.D.: Multiple resources and mental workload. *Hum. Factors* **50**, 449–455 (2008)
2. Young, M.S., Brookhuis, K.A., Wickens, C.D., Hancock, P.A.: State of science: mental workload in ergonomics. *Ergonomics* **58**, 1–17 (2014). (advance online publication)
3. Colle, H.A., Reid, G.B.: Context effects in subjective mental workload ratings. *Hum. Factors* **40**, 591–600 (1998)
4. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) *Human Mental Workload*, pp. 139–184. North-Holland, Amsterdam (1988)
5. Yeh, Y.Y., Wickens, C.D.: Dissociation of performance and subjective measures of workload. *Hum. Factors* **30**, 111–120 (1988)
6. Boles, D.B., Bursk, J.H., Phillips, J.B., Perdelwitz, J.R.: Predicting dual-task performance with the multiple resources questionnaire (MRQ). *Hum. Factors* **49**, 32–45 (2007)
7. Finomore, V.S., Shaw, T.H., Warm, J.S., Matthews, G., Boles, D.B.: Viewing the workload of vigilance through the lenses of the NASA-TLX and the MRQ. *Hum. Factors* **55**, 1044–1063 (2013)
8. Saxby, D.J., Matthews, G., Warm, J.S., Hitchcock, E.M., Neubauer, C.: Active and passive fatigue in simulated driving: discriminating styles of workload regulation and their safety impacts. *J. Exp. Psychol. Appl.* **19**, 287–300 (2013)
9. Hockey, G.R.J.: Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. *Biol. Psychol.* **45**, 73–93 (1997)
10. Young, M.S., Stanton, N.A.: Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Hum. Factors* **44**, 365–375 (2002)
11. Matthews, G., Davies, D.R., Westerman, S.J., Stammers, R.B.: *Human Performance: Cognition, Stress and Individual Differences*. Psychology Press, London (2000)
12. Brookhuis, K.A., de Waard, D.: Monitoring drivers' mental workload in driving simulators using physiological measures. *Accid. Anal. Prev.* **42**, 898–903 (2010)
13. O'Donnell, R.D., Eggemeier, F.T.: Workload assessment methodology. In: Boff, K.R., Thomas, J.P. (eds.) *Handbook of Perception and Human Performance. Cognitive Processes and Performance*, vol. 2, 2nd edn, pp. 1–49. Wiley, Oxford (1986)
14. Matthews, G., Reinerman-Jones, L.E., Barber, D.J., Abich, J.: The psychometrics of mental workload multiple measures are sensitive but divergent. *Hum. Factors* **57**, 125–143 (2014)

15. Reinerman-Jones, L.E., Matthews, G., Barber, D.J., Abich, J.: Psychophysiological metrics for workload are demand-sensitive but multifactorial. *Proceed. Hum. Factors Ergon. Soc. Annu. Meet.* **58**, 974–978 (2014)
16. Reinerman-Jones, L., Guznov, S., Mercado, J., D'Agostino, A.: Developing methodology for experimentation using a nuclear power plant simulator. In: Schmorow, D.D., Fidopiastis, C.M. (eds.) *Foundations of Augmented Cognition*, pp. 181–188. Springer, New York (2013)
17. Loehlin, J.C.: *Latent Variable Models: An Introduction to Factor, Path, and Structural Equation Analysis*. Lawrence Erlbaum, Mahwah (2004)
18. Stephens, C.L., Christie, I.C., Friedman, B.H.: Autonomic specificity of basic emotions: evidence from pattern classification and cluster analysis. *Biol. Psychol.* **84**, 463–473 (2010)
19. Baldwin, C.L., Penaranda, B.N.: Adaptive training using an artificial neural network and EEG metrics for within-and cross-task workload classification. *NeuroImage* **59**, 48–56 (2012)
20. Matthews, G., Warm, J.S., Shaw, T.H., Finomore, V.S.: Predicting battlefield vigilance: a multivariate approach to assessment of attentional resources. *Ergonomics* **57**, 856–875 (2014)
21. Langner, R., Eickhoff, S.B.: Sustaining attention to simple tasks: a meta-analytic review of the neural mechanisms of vigilant attention. *Psychol. Bull.* **139**, 870–900 (2013)
22. Vidulich, M.A., Tsang, P.S.: Mental workload and situation awareness. In: Salvendy, G. (ed.) *Handbook of Human Factors and Ergonomics*, 4th edn, pp. 243–273. Wiley, Hoboken (2012)
23. Mercado, J.E.: *Assessing the effectiveness of workload measures in the nuclear domain*. Doctoral dissertation, University of Central Florida (2014)
24. Matthews, G., Warm, J.S., Reinerman, L.E., Langheim, L., Washburn, D.A., Tripp, L.: Task engagement, cerebral blood flow velocity, and diagnostic monitoring for sustained attention. *J. Exp. Psychol. Appl.* **16**, 187–203 (2010)
25. Stroobant, N., Vingerhoets, G.: Transcranial doppler ultrasonography monitoring of cerebral hemodynamics during performance of cognitive tasks: a review. *Neuropsychol. Rev.* **10**, 213–231 (2000)
26. Wohleber, R.W., Matthews, G., Reinerman-Jones, L.E., Sollins, B., Panganiban, A.R., Scribner, D.R.: Manuscript in preparation (2015)
27. Lin, J., Matthews, G., Wohleber, R., Calhoun, G., Funke, G.: Manuscript in preparation (2015)
28. Matthews, G., Szalma, J., Panganiban, A.R., Neubauer, C., Warm, J.S.: Profiling task stress with the Dundee stress state questionnaire. In: Cavalcanti, L., Azevedo, S. (eds.) *Psychology of Stress: New Research*, pp. 49–91. Nova Science, Hauppauge (2013)
29. Mulder, L.J.M., De Waard, D., Brookhuis, K.A.: Estimating mental effort using heart rate and heart rate variability. In: Stanton, N.A., Hedge, A., Brookhuis, K., Salas, E., Hendrick, H.W. (eds.) *Handbook of Human Factors and Ergonomics Methods*, pp. 201–210. CRC Press, Boca Raton (2004)
30. Thayer, J.F., Åhs, F., Fredrikson, M., Sollers, J.J., Wager, T.D.: A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neurosci. Biobehav. Rev.* **36**, 747–756 (2012)
31. Parasuraman, R., Wickens, C.D.: Humans: still vital after all these years of automation. *Hum. Factors* **50**, 511–520 (2008)