Mental workload and performance measurements in driving task

Citation for published version:
Butmee, T, Lansdown, TC & Walker, GH 2019, Mental workload and performance measurements in driving task: A review literature. in Y Fujita, S Bagnara, R Tartaglia, S Albolino & T Alexander (eds), *Proceedings of the 20th Congress of the International Ergonomics Association: IEA 2018.* Advances in Intelligent Systems and Computing, vol. 823, Springer, pp. 286-294, 20th Congress of the International Ergonomics Association 2018, Florence, Italy, 26/08/18. https://doi.org/10.1007/978-3-319-96074-6_31

Digital Object Identifier (DOI):
10.1007/978-3-319-96074-6_31

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Peer reviewed version

Published In:
Proceedings of the 20th Congress of the International Ergonomics Association

Mental Workload and Performance in the Driving Task: A Literature Review

Totsapon Butmee[1], Terry C. Lansdown[1] , Guy H. Walker[1]

[1]Heriot-Watt University, Edinburgh, United Kingdom

tb19@hw.ac.uk, T.Lansdown@hw.ac.uk, g.h.walker@hw.ac.uk

Abstract. The concept of human mental workload in the field of human factors and psychology has a long history with important applications in the aviation and automotive industries. Mental Workload is a complex concept and it is difficult to define the term. It currently has no universal accepted definition. Mental workload cannot be measured directly, however, it has been shown it that relates to internal resource limitations to accomplish a task, and also impacts multi-dimensionally on other factors. Previously, several studies have been indicated that mental workload relate with operators' performance, task demand and mental resource supply. Extremes (underload or overload) mental workload can degrade operators' performance. Several assessment methods have been proposed for investigating mental workload. They can be performed in experimental or operational settings. There are seven selection criteria to select the most appropriate methods. These include sensitivity, diagnosticity, intrusiveness, implementation requirement, operator acceptance, selectivity and bandwidth and reliability. Dozens of Mental workload measurement techniques have been developed and categorized into three main groups. (i) *Subjective rating,* which were categorized into unidimensional and multidimensional. NASA-TLX, SWAT, RSME and MCH are the famous examples of subjective-based techniques. (ii) *Performance measures* are divided into primary task and secondary task measures. Primary task measures are capable of discriminating the resource competition between individual differences. For example, speed instability, distance headway instability, lateral position from road centreline, lane excursion, time spent out of lane can be widely used to represent the driver primary performances. In secondary task measures are more diagnosticity than primary task measures and subjective measures. Correct response, time response of additional secondary task are a well-known examples of secondary task performance measures in driving research. Additionally, (iii) *Physiological techniques* also have high sensitivity in measurement, but results from these methods can easily are confounded by other external and extraneous interference. Measures of Eye Functions have been frequently used if compared with other Physiological techniques. It can be argued that the combined methods are recommend cooperatively to predict human mental workload.

Keywords: mental workload measurements, performance measurements, driving task

1 Introduction

The Driving task has been defined as numerous highly dynamic tasks occurring in a changing environment [1]. Driving complexity has been reported to depend on several elements including *road design* (city roads vs. rural roads vs. motorways), *traffic flow* (low density vs. high density), and *road layout* (straight vs. curves, even vs. inclined, junction vs. no junction) [2]. It is not purely a physical task (e.g., applying force on steering wheel and pedals), but also a visual and mental one [3]. In the terms of mental task, the driver deals with millions of sensory signals such as visual, verbal and other stimuli. Sometimes, substantial information needs to be processed by the limited resources of the driver. Such situations can impose high demand on the driver's cognitive systems [1]. Mental workload issues have been investigated during the last four decades in various research fields [for example, 1, 4-5]. In driving research, extreme (underload or overload) mental workload may degrade drivers' performance and potentially increase collision rate. For example, it has been reported that many in-vehicle activities (e.g., systems such as navigation system, conversations with a friend on hands-free cell phone, or using a speech-to-text interface for e-mail) have been associated with distraction and information overload during driving [6-7]. Lansdown, Brook-Carter, & Kersloot [8] showed that multiple in-vehicle systems significantly impose high mental workload. On the other hand, engaging for prolonged periods with automated vehicle systems or undertaking monotonously underloading tasks has been also reported to turn the driving task into a vigilance one [1], in which attentional loss and performance decrements are common [9].

Evaluating mental workload is an important component of system design and analysis [10]. Especially in the driving task, mental workload assessment may be highly helpful to reduce the number of road accidents [11]. Mental workload measurement has a long history originating from aviation research. Much driving research has adopted techniques from the previous studies, using the same concepts from aviation to investigate mental workload. The main objectives of this review are to define metal workload in driving task, and summarise the results of recently conducted studies.

2. Definitions

The concept of Human Mental Workload in the field of human factors and psychology has a long history with important applications in the aviation and automotive industries. However, it is difficult to define this term [12]. At the present, there is no clear definition nor universally accepted term for Mental Workload [12-15]

Various views that have been proposed to understand the term mental workload. For example, O'Donnell & Eggemeier [5] defined the term of workload as a 'portion of the operator's limited capacity actually required to perform a particular task'. In more general term, it has been suggested that mental workload can be described by the amount of human 'information processing capacity' which is spent for task performance [16].

An interesting definition, presented by De Waard [1] suggests there are three key terms required to understand mental workload. First, 'task demand' is required to reach the goals and sub-goals by operator's performance. Second, 'workload' is the proportion of the capacity that is allocated for task demand. It is reacted by task demand. Third, 'effort' includes 'state-related effort' which is exerted to maintain an optimal state for task performance, and 'task-related effort' is exerted in the case of controlled information processing. A different perspective by Young et al. [12] explained mental workload as a multidimensional construct, and it is described by 'task' (e.g. demand and performance), 'operator' (e.g. skill and attention) characteristics, and the environmental context. Thus, mental workload is the result of an interaction between task demands individual characteristics, and the environment [3]. A literature review by Xie & Salvendy [15] summarized main tenets of mental workload including: amount of *mental work (effort)* to complete a task (which cannot be detected directly); it further involves the depletion of *individual internal resources* to accomplish the task; and is a *multi-dimensional* factor, i.e. time, mental-effort and psychological-stress loads. In driving research domain, mental workload was specifically defined as the effort to maintain the driving task within a subjective safety zone [17].

As can be seen, mental workload is related to attentional demand on human's information process. In the driving task, for example, drivers usually have higher mental workload under dual- or multiple- task periods. One can compare this with the 'driving only' task where the driver does not need to invest additional attentional resources to meet the task demands. Overload was described as leading to high response times, high error rates, low performance and reduced mental residual capacity to deal with other tasks; while underload may also reduce sustained attention on tasks, increase reaction times and again lower performance. Such significant issues require particular concern, as it may lead to road accidents.

3. Literature Search Methods

The literature search was conducted using Elsevier, Science Direct, Tandfonline and Web of Science. Searches were performed between September 2016 and September 2017 with using one of six following keywords: "mental workload measurements" "performance measurements" "mental workload assessments" "performance assessments" "mental workload evaluations" "performance evaluation" and with one of four following keywords: "driving task" "automobile driving" "car driving" "vehicle driving". In this review, inclusion criteria included the academic articles in peer reviewed journals published in English between January 1969 and September 2017.

4. Mental Workload and Performance Measurements

4.1 Measurement Criteria

Workload measurements should predict an operator's performance. The main reason for mental workload measurement is to quantify the mental cost to perform a task, and thereby to predict operator and systems' performance. However, there are several considerations regarding the selection of workload measurement instruments. Several factors, outlined below, have been proposed for selecting and developing measurement techniques.

There are five selection criteria to determine the most appropriate measurement method. *Sensitivity* is the capability to discriminate the nature of the workload which is imposed by the task(s) demand [5] or reflect changes in workload [1, 18]. It represents the ability of a metric to detect changes in workload. Some studies need to measure the main source of workload such as perceptual versus central processing versus motor resources [5]. So, *diagnosticity* is another important criteria. It can be defined as the capability of technique to identify the specific type of workload or locus of demand [1], e.g., perceptual demand vs. central processing vs. motor resources [5]. Moreover, *intrusiveness* should also be considered, particularly when additional secondary task measures and/or physiological assessment techniques might interfere the primary task. Intrusiveness refers to the degree the method interferes with the performance of a primary task [14]. Ease of use vs. the complexity of the measurement procedure is considered as an *implementation* criteria. For example, the training of operators or the need of specific equipment for data collection and analysis may be referred to as implementation requirements. Particularly, in a real-world situation such as on-road driving condition, the implementation requirements become important criteria [1]. *Operator acceptance* is defined as operator's willingness to follow the methodological instructions and actually utilize a particular technique accurately [5]). It has been suggested that a mental workload measurement method should be simple, both demonstrating an understanding and directness for mental workload measurement [14]. *Selectivity, Bandwidth and reliability* were additionally recommended as additional criteria by Longo [14]. In additional to reliability, a mental workload assessment technique should have 'high sensitivity, high bandwidth, and low intrusiveness' regarding the primary task. Additionally, concurrent and convergent validity should be demonstrable. [14]. Each current measurement technique has its own advantages and disadvantages and thus, some are more appropriate for different contexts. It has been recommended here that using multiple combination mental workload measurement (subjective rating, performance and physiological) techniques to get the most accurate assessment [18] and more comprehensive assessment than using only one technique [7].

4.2 Mental Workload Measurement Techniques

Dozens of mental workload measurement techniques have been developed and can be categorized into three main groups [5, 12, 18-19] including: (i) subjective measurement; (ii) performance-based measurement; or (iii) physiological measurement.

Subjective measures have been developed to assess the perceptions of the operator on the mental workload experienced after a time period or to probe the current level of demand. The main advantages of subjective self-report measures are reducing application cost, high sensitivity to underload and overload situations [14]. Reflective variants (e.g., NASA-TLX [4]) of these techniques do not interfere with primary task [20], while 'probe' measures (e.g., RSME [23]) seem likely to impose some additional demand.

Probe measures are practically vary easy methods of assessing workload [18]. However, they make it difficult to discriminate between, for example, physical- and mental-workload [18], some people cannot distinguish between external demands and actual effort or workload [1, 5]. Subjective rating scales have been further categorized into two groups [14, 18, 20], i) *unidimensional scales*, e.g., Modified Cooper-Harper (MCH, [21]), Subjective Workload Dominance technique (SWORD) [22] and Rating Scale Mental Effort (RSME[23]); and ii) *multidimensional scales*, offering a more diagnostic assessment, e.g., NASA-Task Load Index (NASA-TLX, [4]) or Subjective Workload Assessment Technique (SWAT, [24]). NASA-Task Load Index (NASA-TLX) is well-known and has been frequently used to measure driver mental workload [12]. However, a limitation of 'reflective' subjective methods is the danger of recall bias, such that any workload ratings obtained should be completed as soon as possible after task performance [5].

Performance measures of mental workload are used to quantify how well an operator is undertaking the particular task(s). Performance measures have been categorized further into two categories including *direct* and *indirect measurements* [5, 16, 20]. Direct measurements focus on performance of the main or *primary task* (i.e., in driving avoiding other obstacles while travelling to one's destination, encompassing lateral control and longitudinal control [20]. *Indirect measurements* consider secondary task performance, such as manual responses to a stimulus presented in the visual or auditory field [20] (e.g., the peripheral detection task (PDT, [16]) as a proxy for changes in workload. Both direct and indirect task measures have been used widely in driving literature [18]. Direct (primary) task measurements have been shown to be not sensitive to changes in low workload situations [18]. Further, they are limited with respect to diagnostic capability [5, 25]. Considering secondary task measures, intrusiveness is a key limitation, most techniques require additional operational equipment and/or further operator training.

Physiological measures are based on the concept that increased mental demands lead to observable physical responses from human body [18]. Miller [18] classified five physiological measures: cardiac; respiratory; eye; speech and brain activities. Examples of widely accepted physiological measures are shown below in Table 1. Eye blink measurements (such as blink rate and blink duration) have been used in several driving studies with most accurate for visual workload. Many techniques on cardiac functions measurement such as Heart Rate Variability (HRV) and Inter-Beat-Interval (IBI) have been widely utilized and accepted to measure mental workload [1-2, 27]. To measure brain functions, electroencephalogram (EEG) is frequently used with extremely accurate and reliable [18]. Electromyogram (EMG) signal relates to the force exerted by facial muscle and increased muscular tension level associated with both physical and mental workload. An increment in mental workload or stress represents to an increment in the EMG level [5]. However, a common drawback is that most physiological measures are required special equipment as well as trained operators with technical expertise to utilize these equipment and interpret the data [25]. One concern from physiological assessments is that signals may interfere by other measures, and may therefore provide a misleading indicator.

Table 1. Physiological Measures

Physiological Measures	Examples
1. Eye Behaviour Measurement	Blink rate, blink duration, pupil diameter and Electroocculogram (EOG)
2. Measures of Cardiac Functions	Heart Rate (HR), Heart Rate Variability (HRV), Inter-Beat-Interval (IBI) and Electrocardiogram (EKG)
3. Measures of Brain Functions	Electroencephalogram (EEG)
4. Measures of Muscle Functions	Electromyogram (EMG)
5. Skin	Galvanic Skin Response (GSR)
6. Other	Electrodermal Activity (EDA) and Hormone Levels

Since several techniques have been developed for measuring human mental workload. Miller [18] presented criteria to select the appropriate methods. They consider the Interval of collection, Obtrusiveness, Form of gathering data, Time consideration, Sensitivity, Cost of implementation, and Reliability, see Table 2.

Table 2 Selection of a Mental Workload Technique

Considerations	Type of data	Methods
Interval of collection	Continuous	Physiological Primary performance measures Secondary performance measures
	During	Unidimensional rating
	After	Multidimensional rating
Obtrusiveness	Obtrusive	Brain measures Respiratory measures Secondary performance measures
	Unobtrusive	Subjective Other physiological Primary performance measures
Form of gathering data	Verbal	Unidimensional rating
	Written	Multidimensional rating
	Machine gathered	Physiological Primary performance measures Secondary performance measures
Time consideration	Yes	Unidimensional rating

(only subjective measure)	No	Multidimensional rating
Sensitivity	High	Unidimensional rating Multidimensional rating Brain Measures
	Medium	Secondary performance measures Primary performance measures Cardiac measure Eye measures
	Low	Other physiological
Cost of implementation	High	Brain measures Respiratory measures Eye measures
	Moderate	Multidimensional rating Most primary performance measures Most secondary performance measures Other physiological
	Low	Unidimensional rating
Reliability	High	Unidimensional rating Multidimensional rating Brain Measures Eye activity
	Medium	Secondary performance measures Other physiological
	Low	Primary performance measures

5. Conclusions and Discussion

The aims of this review were to define driving task mental workload and summarize recent studies concerning mental workload measurement criteria and techniques. Several definitions have been presented. However, there were variety views along with different research sectors. Based on these views [1, 3, 12, 15, 17], considering the driving task, it should be defined that mental workload is related to the effort required from the driver to meet the attentional demands on cognitive resources, in their attempts to meet the task demands. However, it has no excessive conclusions to confirm which method is the most appropriate. Evaluating workload is an important component of system design and analysis [10]. A useful mental workload assessment technique should have high reliability, along with 'high sensitivity, a high bandwidth, low intrusiveness on the primary task [1, 5, 18]. Three techniques for mental workload measurement are discussed: subjective, performance and physiological measures. For subjective measures, NASA-TLX is a well-known and widely used technique in driving research. One possible reason was that this multidimensional rating ease of use and can represent the sources of workload [4]. Our review found that most previous studies mentioned performance measurements in their works along with other mental workload measurement techniques. It should be noted that performance-based measurements of driver's mental workload might be the most frequently used methods to indicate driver's mental workload. Several physiological approaches have been used to measure the driver's mental workload. For example, eye tracking. Although several mental workload measurement techniques have been developed more than 40 years, no one has been accepted as the best technique. Thus, more research on driver's mental workload measurements is still required to develop. In the near future, as autonomous systems continue to replace laborious human roles; human factors research will need to investigate the novel workload implications of such sophisticated automobile technologies.

References

1. de Waard, D.: The measurement of drivers' mental workload. Netherlands: Groningen University, Traffic Research Center (1996).
2. Paxion, J., Galy, E., & Berthelon, C.: Mental workload and driving. Frontiers in psychology 5, 1-11 (2014).
3. Marquart, G., Cabrall, C., & de Winter, J.: Review of eye-related measures of drivers' mental workload. Procedia Manufacturing, 3, 2854-2861 (2015).
4. Hart, S. G., & Staveland, L. E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology, 52, 139-183 (1988).
5. O'Donnell, R.D. and Eggemeier, F.T.: Workload assessment methodology In K.R. Boff, L. Kaufman and J.P. Thomas, John Wiley and Sons (Ed.) Handbook of Perceptionand Human Performance. Volume 2. Cognitive Processes and Performance (pp.42-1:42-49). United States : A wiley-interscience publication (1986).
6. Piechullaa, W., Mayserb, C., Gehrkec, H. and Könige, W.: Reducing drivers' mental workload by means of an adaptive man–machine interface, Transportation Research Part F: Traffic Psychology and Behaviour, 6(4), 233–248 (2003).
7. Strayer, D. L., Turrill, J., Cooper, J. M., Coleman, J. R., Medeiros-Ward, N., & Biondi, F.: Assessing cognitive distraction in the automobile. Human factors, 57(8), 1300-1324 (2015).
8. Lansdown, T. C., Brook-Carter, N., & Kersloot, T.: Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications. Ergonomics, 47(1), 91-104 (2004).
9. Körber, M., Cingel, A., Zimmermann, M., & Bengler, K.: Vigilance decrement and passive fatigue caused by monotony in automated driving. Procedia Manufacturing, 3, 2403-2409 (2015).
10. DiDomenico, A., & Nussbaum, M. A.: Interactive effects of physical and mental workload on subjective workload assessment. International journal of industrial ergonomics, 38(11), 977-983 (2008).
11. Marquart, G., Cabrall, C., & de Winter, J.: Review of eye-related measures of drivers' mental workload. Procedia Manufacturing, 3, 2854-2861(2015).
12. Young, M. S., Brookhuis, K. A., Wickens, C. D. and Hancock, P. A.: State of science: mental workload in ergonomics. Ergonomics, 58(1), 1-17 (2015).
13. Cain, B.: A review of the mental workload literature. Defence Research And Development Toronto (Canada) (2007).

14. Longo, L.: A defeasible reasoning framework for human mental workload representation and assessment. Behaviour & Information Technology, 34(8), 758-786 (2015).
15. Xie, B., & Salvendy, G.: Review and reappraisal of modelling and predicting mental workload in single-and multi-task environments. Work & stress, 14(1), 74-99 (2000).
16. Brookhuis, K. A., van Driel, C. J., Hof, T., van Arem, B., & Hoedemaeker, M.: Driving with a congestion assistant; mental workload and acceptance. Applied ergonomics, 40(6), 1019-1025 (2009).
17. Boer, E. R.: Behavioral entropy as a measure of driving performance. In Proceedings of the First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design (pp. 225-229) (2001).
18. Miller, S.: Workload measures. National Advanced Driving Simulator. Iowa City, United States (2001).
19. Cantin, V., Lavallière, M., Simoneaub, M. and Teasdale, N.: Mental workload when driving in a simulator: Effects of age and driving complexity. Accident Analysis & Prevention, 41(4), 763–771 (2009).
20. da Silva, F. P.: Mental Workload, Task Demand and Driving Performance: What Relation?. Procedia-Social and Behavioral Sciences, 162, 310-319 (2014).
21. Cooper, G. E., & Harper Jr, R. P.: The use of pilot rating in the evaluation of aircraft handling qualities (No. AGARD-567). Advisory Group for aerospace research and development Neuilly-Sur-Seine (France) (1969).
22. Vidullch, M. A., Ward, G. F., & Schueren, J.: Using the subjective workload dominance (SWORD) technique for projective workload assessment. Human Factors, 33(6), 677-691 (1991).
23. Zijlstra, F. R. H.: Efficiency in work behaviour: A design approach for modern tools (1993).
24. Reid, G. B., & Nygren, T. E.: The subjective workload assessment technique: A scaling procedure for measuring mental workload. Advances in psychology, 52, 185-218 (1988).
25. Rehmann, A. J.: Handbook of Human Performance Measures and Crew Requirements for Flightdeck Research (No. CSERIAC-ACT-350). CREW SYSTEM ERGONOMICS INFORMATION ANALYSIS CENTER WRIGHT-PATTERSON AFB OH (1995).
26. Ryu, K., & Myung, R.: Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. International Journal of Industrial Ergonomics, 35(11), 991-1009 (2005).
27. Brookhuis, K. A., & de Waard, D.: Monitoring drivers' mental workload in driving simulators using physiological measures. Accident Analysis & Prevention, 42(3), 898-903 (2010).