



# Task Demand Transition Peak Point Effects on Mental Workload Measures Divergence

Enrique Muñoz-de-Escalona<sup>1</sup>(✉), José Juan Cañas<sup>1</sup>, Chiara Leva<sup>2</sup>, and Luca Longo<sup>3</sup>

<sup>1</sup> Mind, Brain and Behaviour Research Centre, University of Granada, Granada, Spain  
{enriquemef, delagado}@ugr.es

<sup>2</sup> School of Environmental Health, Technological University Dublin, Kevin Street, Dublin 2,  
Dublin, Republic of Ireland  
mariachiara.leva@tudublin.ie

<sup>3</sup> School of Computer Science, Technological University Dublin, Dublin, Republic of Ireland  
luca.longo@tudublin.ie

**Abstract.** The capacity to assess and manage mental workload is becoming more and more relevant in the current work environments as it helps to prevent work related accidents and achieve better efficiency and productivity. Mental workload is often measured indirectly by inferring its effects on performance, mental states, and psychophysiological indexes. Since these three main axes should reflect changes in task demands, convergence between measures is expected, however research has found that this convergence is not to be taken for granted as it is not often present. This study aims to explore how the task demand transition peak point may affect in mental workload divergence between measures during task-load changes: some measures might be more sensitive to abrupt changes in task demand than others and this could also be mediated by the task-load baseline. This was tested by manipulating task-load transitions and the point at which the change in the task load magnitude reaches its highest relative peak over time during a monitoring experiment, while psychophysiological (pupil size) and subjective perceptions were collected as indicators of subjects' workload alongside objective indicators of task performance from the simulation. The results showed that performance measure proved to be sensitive to abrupt increases in task demand in every condition whereas our physiological measure was only sensitive to a sudden increase in task-load during low mental workload baseline circumstances. Furthermore, contrary to what expected, subjective ratings of mental workload did not react to abrupt transitions in task-load in every condition but only to an absolute measure of the overall level of task demand.

**Keywords:** Mental workload · Workload measures · Convergence · Divergence · Dissociations · Insensitivities · Task demand transitions · Rates of change · Peak point

## 1 Introduction

Mental workload has been the subject of extensive research in many fields over the last several decades. Despite multiple efforts were made for reaching a consensus regarding

its definition, to date there is no solid agreement on defining a commonly accepted construct [1–4]. However, despite the difficulties in defining mental workload, the need to assess it in many safety critical workplaces is increasing, particularly in the workplace, for interaction [45, 46] and instructional design [47]. Mental workload is closely linked to mental fatigue and performance and so, with human error and work-related accidents [5–10]. On the one hand, a high mental workload situation sustained over time, which is defined as mental overload, increases the chance of making mistakes at work. But also, on the other hand, a very low mental workload can lead to the dangerous and unwanted phenomena defined as “out of the loop” [11, 12]. Hence, assessing mental workload could not only help to predict overload and underload situations and to prevent incidents and fatalities, but also having a good management of mental workload levels would also be advantageous to increase productivity at work [13–15]. There exist several different methodologies for measuring mental workload, but literature research has shown that not all those methodologies do necessarily coincide in providing the same results [16–20]. Empirical research has shown that divergences between measurements exist and, for that reason, the potential causes of these divergences has been the subject of intensive research in recent years [21–26]. There are several different explanations for this lack of convergence between measures. One of these explanations could be the differential effects of “task demand transitions” on mental workload measurements. While a person is performing a task, task demand may increase or decrease. This change in the magnitude of task demand over the duration of a task is what we call a “task demand transition”, which could be smooth or abrupt.

The aim of this paper is to shed some light about mental workload divergences that might be produced by task demand transitions. Particularly, we want to explore how the task demand transition peak point (the moment in which the abrupt increase in task demand occurs) may affect in mental workload divergence between measures during task-load changes. In other words, it may be possible that task demand transition peak point may affect more and in different ways some mental workload measures than others so that it would facilitate the emergence of divergence between them. In Sect. 2 we will briefly review the most important explanations that have been suggested for this lack of convergence between measures, as well as certain key concepts about measuring mental workload. Then, in Sect. 3 we present the design and methodology used to undertake our study. In Sect. 4 we describe the results obtained from the data collection campaign, while in Sect. 5 we discuss our findings and suggest possible future directions in this research area. Finally, in Sect. 6 we provide an overview of the key conclusions derived from the study, as well as its major implications.

## 2 Related Work

### 2.1 Assessing Mental Workload

Assessing mental workload is not a simple task since it is not something that can be measured directly, and it has to be inferred from its different effects on behavior, mental states, and psychophysiological indexes. Mental workload has three well-established components, two of which are observable (performance and physiological) and one that

is a non-observable subjective feeling of workload. In this sense there are three different categories of measures which are commonly used to assess mental workload:

1. **Performance measures:** this category is composed of primary task measures (n° of errors, reaction time) and secondary task measures (choice reaction-time tasks, time estimation, memory-search tasks). The goal is to measure objective task performance indexes with a view to assess the quantity and the quality of performed tasks.
2. **Physiological measures:** it includes physiological responses to mental activation (pupil diameter, heart rate variability (HRV), electroencephalogram (EEG)). The aim is to collect objective physiological data, which reacts to mental workload.
3. **Subjective measures:** comprises self-reported measures (NASA-TLX) which aim is to collect easy and low-cost subjective individual-related mental workload data.

Each different category has its own advantages and disadvantages. Performance measures, for example, have a high diagnostic value, which allows to investigate which cognitive process is mainly involved in performing certain tasks; but it comes with the disadvantage of a high level of intrusiveness, as operators have to perform simultaneous tasks and that can be dangerous in certain contexts [27]. Physiological measures, on the other hand, are very sensitive to mental workload phasic changes and have high internal validity but are also highly intrusive and require special equipment and specific expert knowledge [28]. Finally, subjective self-reporting measures are very low cost and easy to implement but are vulnerable to cognitive biases [29]. In general it is safe to say that the ideal way of assessing mental workload would involve a triangulation of at least one of each different type of measures, so as to obtain a performance measure, a physiological measure and a subjective measure in such a way that each would contribute extra information and reinforce the overall evaluation. In line with this assumption, we could expect that each different measure of mental workload should converge, but the literature shows multiple examples of situation where this assumption could not be verified [16–19, 21, 23].

## 2.2 Convergences and Divergences Between Mental Workload Measurements

Far from what could be expected, the different indexes of mental workload, gathered by using different categories of measures, do not necessarily coincide in providing the same results. Literature research has increasingly shown studies that offer evidence about this lack of correlation among measurements (divergences) [16–19, 21, 23], which could be due to the occurrence of dissociations and/or insensitivities in mental workload measures. Associations occur when mental workload measures tracks task-load change, dissociations take place when a mental workload reflection contradict a change in task demands whereas insensitivities occurs when that workload reflection does not change with a change in task demands. Research has shown that this lack of convergence among different measurements due to dissociations or insensitivities do not only occurs between the three different categories of methodologies, but also between different particular methods within each category. For example, numerous studies have shown proof of divergences among well-established physiological measures which are supposed to reflect the level mental workload experienced and its changes in the same way [20, 21]. For example,

pupil size and blink rate have been proved to be ocular parameters sensitive to task demand changes, but sometimes both can be affected by certain non-workload factors (such as brightness and relative humidity), that would impact on the emergence of dissociations among measures. One can only wonder why do we find divergences between measurements that are supposed to be reflecting the same construct? Researchers have identified multiple possible causes about this phenomenon, which can be related to a lack of specificity or to a lack of diagnosticity. The former would explain divergences by the existence of non-workload factors affecting measurements, while the latter might point to the fact that different measures could be reflecting different aspects of mental workload. Mental workload is a multidimensional construct and these lack of specificity and diagnosticity would be consistent with multi-resources theories [30]. Hancock (2017) has offered many reasons about the occurrence of dissociations and insensitivities among mental workload measures, which he defined as the AID's of workload [22]. Timescale considered between measures is one of the possible causes: it may be possible that some measures reflect task demand changes within seconds while other may show a longer latency. Muñoz-de-Escalona & Cañas (2018) found that divergences among measures are more susceptible to appear after an abrupt change in task demands occurs, due to the emergence of higher latency differences between measurements. They found that the subjective measure of mental workload reacted sooner than pupil size (physiological response) to a high peak in task demands. In other words, divergences between subjective and physiological measures were higher because subjective measure may have reached its maximum value before the physiological measures did [23].

### **2.3 Sensitivity to Rates of Change and Peak Point During Task Demand Transitions**

Hancock (2017) points out in his study that another possible factor underpinning dissociations and insensitivities among assessment methods is the differential sensitivity of mental workload measures to rates of change during task-load transitions [22]. In other words, some measures might be more sensitive to change in task demand than others, so that when abrupt shifts in task-load are experienced, divergences will also appear across different measures due to their different sensitivities towards the rate of change. That is precisely what was found in a previous study showing higher divergences between mental workload measures during variable rate of change conditions than during linear rate of change conditions [23]. According to the study by Muñoz-de-Escalona & Cañas (2019) [23] dissociations in performance and pupil size (physiological measure) appeared after an abrupt change in task demands during increasing task-load conditions, while the subjective measure was not affected by the rate of change in the task-load but rather by the absolute level of task demand experienced. These interesting results revealed that divergences between measures might be higher during increasing variable rate of change conditions. However, this study did not address the possible effects that the task demand transition peak point may have in the emergence of divergence between mental workload measures: it may be possible that some measures are more sensitive to abrupt changes in task-load levels than others and, furthermore, it may also be possible that some measures might be more sensitive to task-load change under certain circumstances. Therefore, this motivated the purpose of this study, to explore the

possible effects that the peak point during task demand transitions may have in mental workload divergences between measures.

### 3 Design and Methodology

The hypothesis of this research is that the effects on mental workload measures will be different depending on when the high shift in task-load arises. When task demands are low, there are many resources available to be used so, if there is a sudden shift in task-load, the amount of mental resources mobilized to cope with the task will be higher than if we depart from a higher task demand situation, since in this latter case, human resources are limited. Thus, in more detail, we hypothesized that we will find higher divergences between mental workload measurements when the high shift in task-load arises in a low task demand situation because physiological measures should react sooner than performance and subjective measures, as they reflect that extra activation needed to cope with the task. In this study we manipulated 2 variables: (1) task-load change and (2) task-load change peak point. This manipulation was tested in a task-battery experiment which participants performed for 20 min to collect (a) perceived mental workload, (b) task performance and (c) pupil size as our three types of complementary mental workload primary measures.

#### 3.1 Materials and Instruments

**The MATB-II Software.** Measurements of task performance were collected through the use of the second version of the Multi-Attribute Task Battery (MATB-II), a computer program designed to evaluate operator performance and workload through means of different tasks similar to those carried out by flight crews, with a user-friendly interface as to allow non-pilot participants to utilize it [31]. MATB-II comes with default event files, which can easily be altered to adapt to the needs or objectives of an experiment. The program records events presented to participants, as well as participants' responses. The MATB-II contains the following four tasks: the system monitoring task (SYSMON), the tracking task (TRACK), the communications task (COMM), and the resource management task (RESMAN) (see Fig. 1).

- 1) The SYSMON task is divided into two sub-tasks: lights and scales. For the lights sub-task, participants are required to respond as fast as possible to a green light that turns off and a red light that turns on, and to turn them back on and off, respectively. For the scale sub-task, participants are required to detect when the lights on four moving scales deviate from their normal position and respond accordingly by clicking on the deviated scale.
- 2) In the TRACK task, during manual mode, participants are required to keep a circular target in the center of an inner box displayed on the program by using a joystick with their left hand (the dominant hand was needed for the use of the mouse). During automatic mode, the circular target will remain in the inner box by itself.
- 3) In the COMM task, an audio message is played with a specific call sign and the participant is required to respond by selecting the appropriate radio and adjusting

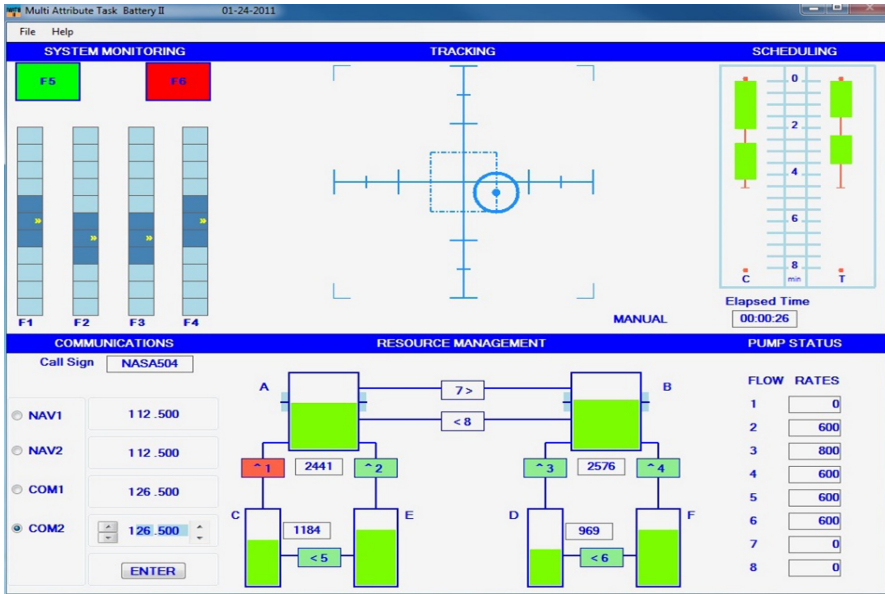


Fig. 1. MATB-II task display. Taken from <https://matb.larc.nasa.gov/>

for the correct frequency, but only if the call sign matches their own (call sign: “NASA504”). No response is required of the participant for messages from other call signs.

- 4) In the RESMAN task, participants are required to maintain the level of fuel in tanks A and B, within  $\pm 500$  units of the initial condition of 2500 units each. In order to maintain this objective, participants must transfer fuel from supply tanks to A and B or transfer fuel between the two tanks.

**The Use of Tobii T120 Eyetracker.** Pupil diameter measurements were obtained using an infrared-based eye tracker system, the Tobii T120 model marketed by Tobii Video System (see Fig. 2). This system is characterized by its high sampling frequency (120 Hz). This equipment is completely non-intrusive, has no visible eye movement monitoring system, and provides high precision and an excellent head compensatory movement mechanism, which ensures high-quality data collection. In addition, a calibration procedure is completed within seconds, and the freedom of movement it offers participants allows them to act naturally in front of the screen, as though it were an ordinary computer display. Thus, the equipment allows for natural conditions in which to measure eye-tracking data [32].

**The Use of the Instantaneous Self-assessment Scale.** We employed an easy and intuitive instant subjective workload scale called instantaneous self-assessment (ISA), which provides momentary subjective ratings of perceived mental workload during task performance (see Fig. 3). ISA has been used extensively in numerous domains, including during ATC tasks. Participants write down how much mental workload they currently



**Fig. 2.** Tobii T120 Eyetracker system

experience on a scale ranging from 1 (no mental workload) to 9 (maximum mental workload), presented from left to right in ascending order of mental workload experienced. This broad range would help us to obtain a good granularity of collected data. Participants were taught to use the scale just before beginning the experimental stage. While the method is relatively obtrusive, it was considered the least intrusive of the available online workload assessment techniques [33, 34].

Participant N°:                      Exp. Condition:                      Age & Gender:

**ISA SCALE**

Please, answer on a scale from 1 to 5, meaning:  
1 (Very low Mental Workload) & 10 (Very high MW).

2.5 MIN----- 1 2 3 4 5 6 7 8 9 10

5 MIN----- 1 2 3 4 5 6 7 8 9 10

**Fig. 3.** Instantaneous self-assessment scale

### 3.2 Participants

The experiment was run with 45 undergraduate and postgraduate students from two Irish universities. Participants’ ages ranged from 20 to 42, with an average of 23.4 and a standard deviation of 4.75. A total of 26 women and 19 men participated. Participation was voluntary and the participants were recruited by providing a seminar about mental workload and its measurements during normal lessons related to Human factors topics. The requirements for participation included (1) not being familiar with the MATB-II program, (2) English as a native language, and (3) visual acuity or correction of visual impairment with contact lenses, as glasses impair the utilized eye-tracking device from collecting data. The voluntary participation was rewarded with extra credit.

### 3.3 Procedure

Participants went through an experimental session consisting of two phases:

1. **Training stage:** a 30 min training session allowed each participant to familiarize themselves with the program so that they could carry out the tasks securely during the data collection stage. The procedure was conducted as follows: upon entering the lab and after filling out the informed consent form, the participant was asked to read the MATB-II instruction manual. The researcher then sat down with the participant to allow for questions and resolve any doubts on how to use the program. Afterward, on a computer monitor, participants were presented with each MATB-II task separately and were first given a demonstration as to how to execute the task. Following this they were given some time (3 min or more if needed) to try and perform the task themselves. Participants would then evaluate task difficulty on a scale ranging from 1 (very easy) to 9 (very difficult). The average difficulties evaluated by the participants for each task of the MATB-II are reported in Table 1.

**Table 1.** Task Difficulty Level Average Assessed by Participants.

Task	Difficulty Average
TRACK	4.16 ( $\sigma$ 1.69)
SYSMON	2.71 ( $\sigma$ 1.39)
COMM	3.33 ( $\sigma$ 1.97)
RESMAN	5.11 ( $\sigma$ 1.95)

The participants were always free to consult the manual and ask the researcher questions during the training stage in case of doubts or uncertainties. Once each participant had completed all four tasks and resolved all doubts, they were ready for the data collection stage, which followed immediately afterwards. During the training stage, participants could work in one room equipped for training with the MATB-II software, and no performance shaping factors was recorded in relation to the environmental conditions as we ensure to have similar average luminosity and comfort for temperature and seating arrangements.

2. **Data collection stage:** the data collection stage lasted a period of 20 min for each participant and it was divided into four intervals of 5 min. In each interval participants had to perform one combination of tasks according to the assigned experimental conditions, while task performance, perceived mental workload, and pupil diameter were recorded. The participants were instructed to fill in the ISA scale every 2,5 min when a scheduled alarm sounded. Prior to the start of the MATB-II tasks, the eye-tracker system was calibrated for each participant while he or she was instructed to keep head and body movements to a minimum. During the data collection stage ensuring conformity with standard room conditions was essential to avoid external performance shaping factors to interfere with the results of the experiment. Thus, the testing rooms were temperature controlled to 21 °C, and lighting conditions (the



main extraneous variable in pupil diameter measurement) were kept constant with artificial lighting; there was no natural light in the rooms. Moreover, participants always sat in the same place, a comfortable chair spaced 60 cm from the eye-tracker system.

This study was carried out in accordance with the recommendations of the local ethical guidelines of the committee of the University of Granada institution: Comité de Ética de Investigación Humana. The protocol was approved by the Comité de Ética de Investigación Humana under the code: 779/CEIH/2019. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

### 3.4 Variables

**Independent Variables** - We manipulated 2 independent variables:

1. **Task-load change:** this variable describes how task demand changes over time. Experimental conditions followed an increasing task demand pattern with a variable rate of change over time.
2. **Task-load rate of change peak point:** this variable describes when the participants experienced the most abrupt change in task demands.

First, we manipulated task-load by modifying the number and the combination of MATB-II active tasks that participants had to perform over time. We created six combination of tasks. Each combination of task made up a condition of increasing difficulty. Table 2 reports the task combinations considered. Task-load increases as the number of active task increases, but also as the relative difficulty of active tasks increases as well. Therefore, since task-load in each individual task was different, the overall task-load in each task combination allowed us to manipulate the task-load change peak point during the experimental session. For example, task combinations 2 and 3 both include 2 tasks, but task-load is higher in task combination 3 as the RESMAN task is more difficult than SYSMON task. Similarly, this is also true for task combinations 4 and 5 (both includes three tasks but task combination 5 is more difficult because it included the RESMAN tasks instead of the COMM task).

**Table 2.** Possible Sets of Task Combinations.

Task Combination	MATB-II Active tasks
1	TRACK
2	TRACK + SYSMON
3	TRACK + RESMAN
4	TRACK + SYSMON + COMM
5	TRACK + SYSMON + RESMAN
6	TRACK + SYSMON + COMM + RESSMAN

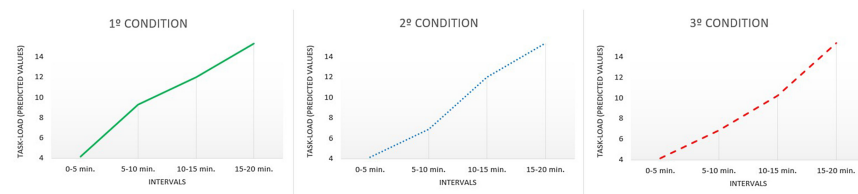
Then, to increase the task-load along the four intervals of five minutes during the experimental session, we created three experimental conditions each one corresponding

to a different choice for the temporary moment in which the magnitude of task-load changes reaching its highest relative peak over time:

- in combination (1) the steepest change in task load is given during early stages of the overall task-load condition
- in combination (2) the steepest change in task load is given during medium stages of the overall task-load condition
- in combination (3) the steepest change in task load is given during the later stages of the overall task-load condition.

The three experimental conditions were:

1. **Condition 1: Highest rate of change during early stages of overall task load.** Task-load increased every 5 min, with a variable rate of change. Participants had to perform the following task combination: 1, 3, 5, and 6. The initial task-load in the first interval (task combination 1) was followed by an abrupt increase in the second interval (task combination 3) and from there, there was a decrease in the rate of change from task combination 3 moving onto 5 and then 6.
2. **Condition 2: Highest rate of change during medium stages of overall task load.** Task-load increased every 5 min, with a variable rate of change. Participants had to perform the following task combination: 1, 2, 5, and 6. The initial task-load of the first interval (task set 1) is followed by a slight task-load increase in the second interval (task set 2), then we find an abrupt increase in task-load rate of change where the participants were moved to task combination 5 and finally a slight decrease in the rate of change for task combination 6.
3. **Condition 3: Highest rate of change during later stages of overall task load.** Task-load increased every 5 min, with a variable rate of change. Participants had to perform the following task combination: 1, 2, 4, and 6. The initial task-load of the first interval (task set 1) is followed by a slight task-load increase moving into task set 2, followed by another slight change (moving into task combination 4) and finally we find an abrupt increase in task-load rate of change moving into task combination 6 (Fig. 4).



**Fig. 4.** Task-load rate of change evolution (current task-load) for the three experimental conditions.

By comparing these three conditions, we were able to measure the effect of task-load rate of change. A  $3 \times 4$  mixed factorial experimental design was devised with

2 independent variables. One variable was task-load rate of change with 3 different levels that were manipulated between groups. The other variable was the interval when task-load changed, this variable had 4 levels and was manipulated within-subjects.

### **Dependent Variables**

*Performance.* MATB-II records many indicators of participants' performance: root mean square deviation (RMSD) for the TRACK task, number of correct and incorrect responses for the SYSMON and COMM tasks, and the arithmetic mean of tanks "A-2500" and "B-2500" in absolute values for the RESMAN task. However, for the purposes of this experiment we will only consider the RMSD performance indicator, as it is the only task present during all task combinations in every experimental condition, allowing us to compare participants' performances through and between conditions. The SRMSD performance indicator reflects the distance of the circle to the target point, so that a higher score on this variable reflects a negative performance.

*Pupil Size.* Mental workload can be revealed by several physiological indexes such as EEG, HVR, and several ocular metrics. We decided to use pupil diameter as our physiological mental workload indicator, as it has been verified in the literature as an objective indicator for mental workload [35–43] and minimizes intrusiveness. While the eye-tracking system used allows a continuous sampling rate recording at 120 Hz, a total of 8 intervals lasting 2.5 min was set, so as to obtain 2 measures in each of the three intervals of different overall task-load levels. Since expressing pupil size in absolute values has the disadvantage of being affected by slow random fluctuations in pupil size (source of noise), we followed the recommendations provided by Sebastiaan Mathôt [44] regarding the baseline correction of pupil-size data. To do this, for every participant, we took his/her average pupil size during the session as a whole as a reference, which was then subtracted from the obtained value in each of the 8 intervals, thereby giving a differential standardized value that allowed us to reduce noise in the collected data. Analyses were carried out for the average of both the left and right pupils. A negative value meant that the pupil was contracting while a positive value meant that it was dilating.

*Subjective Mental Workload.* Traditional offline subjective workload assessment tools, such as the NASA Task load Index (NASA-TLX), do not allow researchers to obtain continuous subjective ratings from participants. In order to facilitate and establish comparisons between mental workload measures, it was necessary to obtain the subjective momentary ratings continuously throughout the experimental session. With this goal, we used the ISA, which is an online subjective workload scale created for this purpose. Ratings were obtained at 2.5-minute intervals throughout the 20 min of the experiment duration, obtaining a total of 8 subjective mental workload ratings (2 measures in each of the three intervals of different overall task-load levels).

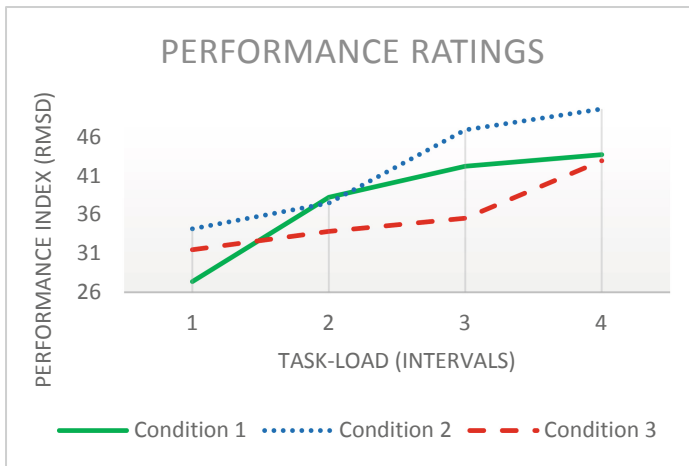
*Synchronization of Measures.* Performance, pupil size, and subjective measures were obtained continuously throughout the experimental session. Synchronization between measures was simple, as the eyetracker and MATB-II performance log files began to record data simultaneously at the start of the experimental session. The scheduled alarm (every 2.5 min) was also synchronized by the experimenter, as it was simultaneously

activated with the MATB-II software. This allowed the ISA scale to be synchronized with the performance and pupil size measures as well.

## 4 Results

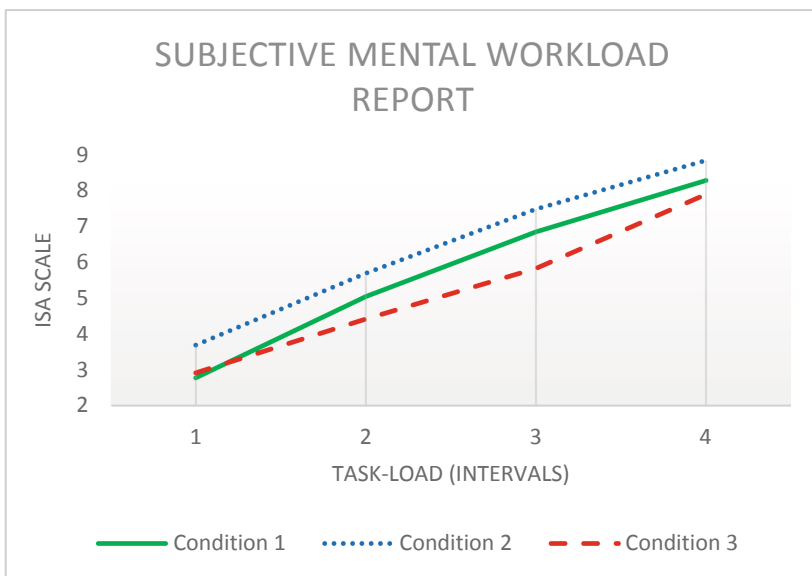
To analyze collected results, we performed three one-way within-subjects ANOVA, one for each different workload measure. First, the ANOVA analysis for our performance measure revealed a significant task-load change main effect  $F(3,126) = 42.53$ ,  $MSe = 40.89$ ,  $p < .01$ , which means that participants' performance was impaired with the increase in task-load through intervals. On the other hand, although the main effect of the experimental condition was not significant  $F(2,42) = 1.33$ ,  $MSe = 427.5$ ,  $p > .05$ , the interaction effect task-load level by experimental condition was found to be significant  $F(6,126) = 2.99$ ,  $MSe = 40.89$ ,  $p < .01$ . This interaction was due to the different evolution of performance impairment through intervals depending on different experimental condition: in experimental condition 1, the greatest difficulties in performance normally occurred between intervals 1 to 2; in experimental condition 2 between intervals 2 to 3 and in experimental condition 3 between intervals 3 to 4. In order to confirm the above statements, we performed 3 mixed factorial ANOVAs, with experimental condition as between group factor, and a one pair of intervals (1 versus 2, 2 versus 3, and 3 versus 4) as the within-subject factor. The first partial ANOVA (int. 1 versus 2) showed a significant task-load change main effect  $F(1,42) = 16.14$ ,  $MSe = 41.87$ ,  $p < .01$ , a non-significant experimental condition main effect  $F(2,42) = .65$ ,  $MSe = 142.34$ ,  $p > .05$ , but also a significant interaction effect of task-load by experimental condition  $F(2,42) = 3.87$ ,  $MSe = 41.87$ ,  $p < .05$ . This interaction effect was explained by the abrupt impairment in task performance in experimental condition 1 that occurred in the second interval. This impairment evolved in a smoother way in experimental conditions 2 and 3. The second partial ANOVA (int. 2 versus 3), revealed a significant task-load main effect  $F(1,42) = 16.19$ ,  $MSe = 34.93$ ,  $p < .01$ , as well as a non-significant experimental condition main effect  $F(2,42) = 1.70$ ,  $MSe = 270.50$ ,  $p > .05$ , but a significant interaction effect task-load change by experimental condition  $F(2,42) = 3.26$ ,  $MSe = 34.93$ ,  $p < .05$ . These results were probably caused by the issues experienced by participants with performance recorded for experimental condition 2. Participants were better able to manage their performance in experimental conditions 1 and 3. In the third partial ANOVA (int. 3 versus 4), task-load change main effect was found to be significant  $F(1,42) = 13.73$ ,  $MSe = 24.13$ ,  $p < .01$ , experimental condition main effect was not found to be significant  $F(2,42) = 3.10$ ,  $MSe = 24.13$ ,  $p > .05$ , but again we found an interaction effect task-load change by experimental condition  $F(2,42) = 3.10$ ,  $MSe = 24.13$ ,  $p = .05$ . In this case participants experienced more difficulty with task load change within experimental condition 3, while in conditions 1 and 2 the impairment was smoother. Therefore, these data confirmed that the greatest performance impairment occurred when there was an abrupt increment in task-load (Fig. 5).

With respect to subjective perceptions, we found a linear increase in the subjective measure of mental workload in every experimental condition. ANOVAs' analysis revealed a significant main effect of task-load  $F(3,126) = 242.5$ ;  $MSe = .92$ ,  $p < .01$ .



**Fig. 5.** Participants' performance during task development.

However, the overall task load effect of the experimental condition was not found to be significant  $F(2,42) = 3.19$ ;  $MSe = 6.33$ ,  $p > .05$ , nor the interaction effect of task-load level by experimental condition,  $F(6,126) = 1.15$ ;  $MSe = .92$ ,  $p > .05$ . Trend analysis confirmed this linear trend of task-load change effect  $F(1,42) = 339.26$ ;  $MSe = 1.97$ ,  $p < .01$  which means that there existed the same lineal increase in mental workload subjective perceptions through intervals in every experimental condition (Fig. 6).

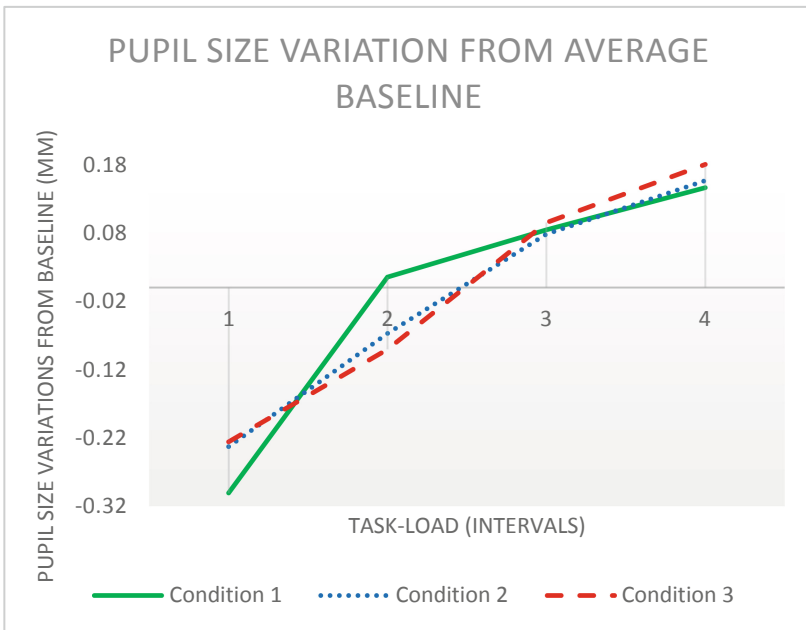


**Fig. 6.** Participants' subjective mental workload ratings during task development.

Finally, regarding our physiological variable, we found a significant effect for the overall task load  $F(3,126) = 198.81$ ,  $MSe = .007$ ,  $p < .01$ . And, although the main effect of the experimental condition was not found to be significant  $F(2,42) = .90$ ,  $MSe = .001$ ,  $p > .05$ , the interaction effect of task-load change by experimental condition was found to be significant,  $F(6,126) = 3.09$ ,  $MSe = .007$ ,  $p < .01$ . As we can see in the graph, pupil size increased throughout the various intervals in every experimental condition. However, whereas in experimental condition 1 we can observe an abrupt increase in pupil size from interval 1 to 2, and then a slight increase from interval 2 to 4, in experimental conditions 2 and 3 we did not find any sudden abrupt dilation change within intervals. These observations were supported by partial ANOVA analyses. A 3-mixed factorial ANOVAs was performed, with experimental condition as between group factor, and one pair of intervals as the within-subject factor. The first partial ANOVA (int. 1 versus 2) revealed a significant task-load change main effect  $F(1,42) = 176.61$ ,  $MSe = .005$ ,  $p < .01$ , a non-significant experimental condition main effect  $F(2,42) = .19$ ,  $MSe = .009$ ,  $p > .05$ , but also a significant interaction effect of task-load by experimental condition  $F(2,42) = 11.77$ ,  $MSe = .005$ ,  $p < .01$ . This interaction effect could be explained by the abrupt pupil size increase in experimental condition 1, while in the other groups, pupil size increased in a smoother and parallel way. In the second partial ANOVA (int. 2 versus 3), we found a significant task-load main effect  $F(1,42) = 59.29$ ,  $MSe = .007$ ,  $p < .01$ , a significant experimental condition main effect  $F(2,42) = 3.46$ ,  $MSe = .007$ ,  $p < .05$  and a significant interaction effect task-load x experimental condition  $F(2,42) = 3.46$ ,  $MSe = .007$ ,  $p < .05$ . This interaction effect was due again to the different course of pupil size in experimental condition 1 with regard to experimental conditions 2 and 3, that is, pupil size increment is much slower in condition 1 than in conditions 2 and 3 which both follow a similar course again. Finally, the third partial ANOVA (int. 3 versus 4) showed a significant task-load change main effect  $F(1,42) = 71.17$ ,  $MSe = .002$ ,  $p < .01$ , a non-significant experimental condition main effect  $F(2,42) = .55$ ,  $MSe = .002$ ,  $p > .05$ , but in this case a non-significant interaction effect of task-load by experimental condition  $F(2,42) = .55$ ,  $MSe = .002$ ,  $p > .05$ . This non-significant interaction could be explained by the absence of significant differences between experimental conditions regarding pupil dilation (Fig. 7).

## 5 Discussion

In summary the results of the experiment showed that performance was negatively affected by overall task-load increased in every experimental condition, but we found that the participants experienced their greatest difficulty with performance in correspondence with the highest rate of change for the task load in each experimental condition. On the other hand, our physiological index, pupil size, also increased as task-load increased in every experimental condition but was not significantly affected by the rate of change in task load with the exception of the results recorded for experimental condition 1, where the highest dilation change matches the task-load change peak point. Finally, regarding subjective perception of mental workload, we did not find any differences between experimental conditions, as it increases linearly through intervals in every condition.



**Fig. 7.** Participants' pupil size variation ratings from the average baseline during task development.

Hence, considering our results, we found a triple association, which means that the three measures of mental workload reflected this construct successfully. However, it is also true that we found differences in the way each measure evolved through intervals:

- Our **performance index** reflected successfully task demand increase through intervals, but it also appeared to be sensitive to abrupt changes in task demands independently of the timing when this abrupt change took place. In other words, our performance measure seemed to be sensitive to abrupt changes at all times, that is, regardless of the current level of task demand.
- Our **subjective index**, on the other hand, seemed to be sensitive to task-load change, but was not sensitive to task-load rate of change regardless of the moment in which the abrupt change in task demand took place. We can appreciate in our results that subjective perceptions of mental workload increased linearly in every experimental condition with no abrupt changes through intervals. In other words, our subjective measure of mental workload seemed to be more sensitive to the absolute level of task demand than to task-load rate of change.
- Regarding our **physiological index**, our results showed that it was sensitive to task-load change as pupil size dilated through intervals. However, the key finding of this research is that pupil size index was also sensitive to abrupt changes in task demand but only when the abrupt change took place in a low task demand situation (experimental condition 1): we can see that during the “early task-load change peak condition” pupil

size greatly increased from interval 1 to 2, that is, when the abrupt change in task-load occurs.

Considering the evolution of the different measures explained above, and in line with what Muñoz-de-Escalona & Cañas (2019) found in their previous research, we have seen that some measures are more sensitive to the absolute level of task demands regardless of task-load rate of change and some others, are more sensitive to abrupt changes in task-load levels. However, it was also found that pupil dilation might be able to reflect abrupt changes in task demand under low task-load circumstances. Therefore, this results confirmed our hypothesis that there would be higher divergences between mental workload measures when an abrupt increase in task-load occurs during low mental workload situations: pupil size measure overreacted to that sudden change in task-load, while performance impairment also increased but not that sharply and subjective perceptions did it in a much smoother way. However, when the abrupt increase in task demand occurred with medium and high mental workload values, pupil size did not increase abruptly, while performance kept showing the effect of the abrupt change in task demands. And subjective perception of mental workload increased as the absolute value of task load increased but was not affected by any abrupt change in task demand. These results can be explained under resource theories. When mental workload is low, there are plenty of mental resources available to provide to cope with an abrupt increase of task demand. However, when workload level is higher, the absolute pool of mental resources could be running out and there would not be many more resources available to be assigned to the task to cope with the abrupt change in task demands. Therefore, human organism can mobilize much more resources in low task demand situations than in higher task demand situations in which the organism will try to save resources as they are more limited and can be depleted. This effect of the abrupt change in task demands is well reflected by the performance of the participants in the task. However, an additional factor we must consider is the fact that the pupil is a muscle and it has a dilation range, so that, when it approximates its limit it will not be able to continue increasing and that would also contribute to the way it is able or not to reflect changes in mental workload. Hence, this could be a contributing reason why pupil size only was able to reflect the abrupt change in task demand during low mental workload condition.

Finally, we again observed that the subjective feeling of mental workload is independent of some factors that affect performance and psychophysiological parameters of mental resources. People are more sensitive to absolute increases of task demand, as shown in other studies of mental workload and fatigue [24, 25]. This experiment however has some limitations that should be noted as the study was performed with a restricted number of students and we consider that this results should be validated with a more representative and numerous sample of the population in order to increase external validation. On the other hand, it has been repeatedly reported in the literature that divergences have been found not only between the different types of methodologies for measuring mental workload, but also within different measures of each methodology. It would have been interesting, for example, to register multiple physiological measures, which reflects mental activation, in order to analyze convergences and divergences within measures. In addition, it would have been interesting to gather more neurophysiological signals



to support obtained findings about activation changes through conditions. Mental workload divergence between measures is a problem which must be addressed with further research, as the factors affecting this phenomenon are multiple and difficult to interpret. Future research could also consider age effects in workload measure divergences which given the nature of the current sample was not significantly observed.

## 6 Conclusions

Mental workload assessment has been the subject of numerous studies in the last several years. Quantifying mental workload not only helps to prevent work-related accidents, but also increases productivity at work, so there is an increasing need in our modern society to measure a construct that even, to date, is very difficult to be defined. There exist three main axes for measuring mental workload: performance, physiological and subjective primary measures. Although it is natural to think that mental workload measures of the same person in the same situation should converge, previous studies have proven that divergences between mental workload measures are very frequent. This study aimed to shed some light about one possible cause for mental workload divergences. Particularly, we wanted to explore how abrupt increases in task demand would affect mental workload measures divergences, depending on the baseline task-load situation. Our findings suggest that although the three considered measures of mental workload were sensitive to the construct, there are differences that we must consider between the way each different measure reacts to sudden increases in mental workload. Firstly, performance measure proved to be sensitive to abrupt increases in task demand in every condition whereas our physiological measure (pupil size) was only sensitive to a sudden increase in task-load during low mental workload baseline circumstances. While subjective ratings of mental workload did not react to abrupt transitions in task-load in any experimental condition but only to the absolute increase in task load.

An important implication about this finding is that we must be cautious in selecting the methodology for assessing mental workload: the decision should be based on our particular goals, as each measure may be more suitable to represent different aspects of mental workload. For example, if we need to detect sudden changes in mental workload in every condition, it could be a better idea to trust performance index, however, if the goal is to detect sudden increases in mental workload during low mental workload sustained situations (certain train drivers) it would be a better idea to trust pupil size, as it seems to be very sensitive to abrupt increase in task-load in this particular situation. Finally, subjective perceptions of mental workload is sufficiently accurate to give an overall idea about the state in which the person finds himself/herself. In any case it is to be assumed that the best way to assess mental workload involves a triangulation of different types of measures, one for each different kind of methodologies, as every different index have its pros and cons and they all contribute relevant information. Last but not least we must be cautious about these findings, as it was obtained for a limited sample therefore further research is needed to continue clarifying the topic of mental workload assessment.

## References

1. Babiloni, F.: Mental workload monitoring: new perspectives from neuroscience. In: Longo, L., Leva, M.C. (eds.) *H-WORKLOAD 2019*. CCIS, vol. 1107, pp. 3–19. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-32423-0\\_1](https://doi.org/10.1007/978-3-030-32423-0_1)
2. Cain, B.: A review of the mental workload literature. Defence Research and Development Toronto (Canada) (2007)
3. Meshkati, N., Hancock, P.A. (eds.) *Human mental workload*, vol. 52. Elsevier (2013)
4. Moray, N. (ed.) *Mental Workload: Its Theory and Measurement*, vol. 8. Springer, New York (2011). <http://doi.org/10.1007/978-1-4757-0884-4>
5. Josten, E.J., Ng-A-Tham, J.E., Thierry, H.: The effects of extended workdays on fatigue, health, performance and satisfaction in nursing. *J. Adv. Nursing* **44**(6), 643–652 (2003). <https://doi.org/10.1046/j.0309-2402.2003.02854.x>
6. Taylor, A.H., Dorn, L.: Stress, fatigue, health, and risk of road traffic accidents among professional drivers: the contribution of physical inactivity. *Annu. Rev. Public Health* **27**, 371–391 (2006). <https://doi.org/10.1146/annurev.publhealth.27.021405.102117>
7. Lilley, R., Feyer, A.M., Kirk, P., Gander, P.: A survey of forest workers in New Zealand: do hours of work, rest, and recovery play a role in accidents and injury? *J. Safety Res.* **33**(1), 53–71 (2002)
8. Sarsangi, V., Salehiniya, H., Hannani, M., Marzaleh, M.A., Abadi, Y.S., Honarjoo, F., Derakhshanjazari, M.: Assessment of workload effect on nursing occupational accidents in hospitals of Kashan Iran. *Biomed. Res. Therapy* **4**(8), 1527–1540 (2017)
9. Rodgers, S.H.: An ergonomic approach to analyzing workplace accidents. *Appl. occupat. Environ. Hygiene* **15**(7), 529–534 (2000)
10. Kirschenbaum, A., Oigenblick, L., Goldberg, A.I.: Well being, work environment and work accidents. *Soc. Sci. Med.* **50**(5), 631–639 (2000)
11. Endsley, M.R., Kiris, E.O.: The out-of-the-loop performance problem and level of control in automation. *Hum. Factors* **37**(2), 381–394 (1995)
12. Young, M.S., Stanton, N.A.: Attention and automation: new perspectives on mental underload and performance. *Theor. Issues Ergon. Sci.* **3**(2), 178–194 (2002). <https://doi.org.ezproxy.un.unimaas.nl/10.1080/14639220210123789>
13. Priya, D.S., Johnson, P., Padmavathi, R., Subhashini, A.S., Ayyappan, R., Ayyappan, M.: Evaluation of the relationship between workload and work capacity in petrochemical and tannery workers—a pilot study. *Life Sci. Med. Res.* **19**(1), 2–12 (2010)
14. Smith, A.P., Smith, H.N.: Workload, fatigue and performance in the rail industry. In: Longo, L., Leva, M.C. (eds.) *H-WORKLOAD 2017*. CCIS, vol. 726, pp. 251–263. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-61061-0\\_17](https://doi.org/10.1007/978-3-319-61061-0_17)
15. Fan, J., Smith, Andrew P.: The impact of workload and fatigue on performance. In: Longo, L., Leva, M. (eds.) *H-WORKLOAD 2017*. CCIS, vol. 726, pp. 90–105. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-61061-0\\_6](https://doi.org/10.1007/978-3-319-61061-0_6)
16. Yeh, Y.Y., Wickens, C.D.: Dissociation of performance and subjective measures of workload. *Hum. Factors* **30**(1), 111–120 (1988)
17. Yeh, Y.H., Wickens, C.D.: The dissociation of subjective measures of mental workload and performance (final report). (No. NASA-CR-176609; NAS 1.26:176609; EPL-84-2/NASA-84-2) (1984)
18. Casper, P.A. Dissociations among measures of mental workload: Effects of experimenter-induced inadequacy (1988)
19. Horrey, W.J., Lesch, M.F., Garabet, A.: Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. *J. Safety Res.* **40**(1), 7–12 (2009)

20. Kamzanova, A.T., Kustubayeva, A.M., Matthews, G.: Use of EEG workload indices for diagnostic monitoring of vigilance decrement. *Hum. Factors* **56**(6), 1136–1149 (2014)
21. Matthews, G., Reinerman-Jones, L.E., Barber, D.J., Abich IV, J.: The psychometrics of mental workload: multiple measures are sensitive but divergent. *Hum. Factors* **57**(1), 125–143 (2015)
22. Hancock, P.A.: Whither workload? mapping a path for its future development. In: Longo, Luca, Leva, M.C. (eds.) *H-WORKLOAD 2017*. CCIS, vol. 726, pp. 3–17. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-61061-0\\_1](https://doi.org/10.1007/978-3-319-61061-0_1)
23. Muñoz-de-Escalona, E., Cañas, J.J.: Latency differences between mental workload measures in detecting workload changes. In: Longo, Luca, Leva, M.C. (eds.) *H-WORKLOAD 2018*. CCIS, vol. 1012, pp. 131–146. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-14273-5\\_8](https://doi.org/10.1007/978-3-030-14273-5_8)
24. Muñoz-de-Escalona, E., Cañas, J.J., van Nes, J.: Task demand transition rates of change effects on mental workload measures divergence. In: Longo, L., Leva, M.C. (eds.) *H-WORKLOAD 2019*. CCIS, vol. 1107, pp. 48–65. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-32423-0\\_4](https://doi.org/10.1007/978-3-030-32423-0_4)
25. Muñoz-de-Escalona, E., Cañas, J.J., Noriega, P.: Inconsistencies between mental fatigue measures under compensatory control theories. *Psicológica Journal*, 1 (ahead-of-print) (2020)
26. Hancock, P.A., Matthews, G.: Workload and performance: associations, insensitivities, and dissociations. *Hum. Factors* **61**(3), 374–392 (2019)
27. Canas, J., Quesada, J., Antolí, A., Fajardo, I.: Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics* **46**(5), 482–501 (2003)
28. Ahlstrom, U., Friedman-Berg, F.J.: Using eye movement activity as a correlate of cognitive workload. *Int. J. Ind. Ergonomics* **36**(7), 623–636 (2006)
29. Bertrand, M., Mullainathan, S.: Do people mean what they say? Implications for subjective survey data. *American Econ. Rev.* **91**(2), 67–72 (2001)
30. Wickens, C.D.: Multiple resources and mental workload. *Human Factors*, **50**(3), 449–455 (2008). <http://doi.org/10.1518%2F001872008X288394>
31. Santiago-Espada, Y., Myer, R.R., Latorella, K.A., Comstock Jr, J.R.: The multi-attribute task battery ii (math-ii) software for human performance and workload research: A user’s guide (2011)
32. Lee, J., Ahn, J.H.: Attention to banner ads and their effectiveness: an eye-tracking approach. *Int. J. Electron. Commerce* **17**(1), 119–137 (2012). <https://doi.org/10.2753/JEC1086-4415170105>
33. Brennan, S.D: An experimental report on rating scale descriptor sets for the instantaneous self assessment (ISA) recorder. DRA Technical Memorandum (CAD5) 92017, DRA Maritime Command and Control Division, Portsmouth (1992)
34. Jordan, C.S.: Experimental study of the effect of an instantaneous self assessment workload recorder on task performance. DRA Technical Memorandum (CAD5) 92011. DRA Maritime Command Control Division, Portsmouth (1992)
35. Matthews, G., Middleton, W., Gilmartin, B.Y., Bullimore, M.A.: Pupillary diameter and cognitive and cognitive load. *J. Psychophysiol.* **5**, 265–271 (1991)
36. Backs, R.W.Y., Walrath, L.C.: Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Appl. Ergon.* **23**, 243–254 (1992). [https://doi.org/10.1016/0003-6870\(92\)90152-1](https://doi.org/10.1016/0003-6870(92)90152-1)
37. Hyönä, J., Tammola, J., Alaja, A.: Pupil dilation as a measure of processing load in simultaneous interpreting and other language tasks. *Q. J. Exp. Psychol.* **48**, 598–612 (1995). <https://doi.org/10.1080/14640749508401407>
38. Granholm, E., Asarnow, R.F., Sarkin, A.J., Dykes, K.L.: Pupillary responses index cognitive resource limitations. *Psychophysiology*, **33**, 457–461 (1996). <http://doi.org/10.1111/j.1469-8986.1996.tb01071.x>

39. Iqbal, S.T., Zheng, X.S., Bailey, B.P.: Task evoked pupillary response to mental workload in human-computer interaction. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, pp. 1477–1480. ACM, New York (2004). <https://doi.org/10.1145/985921.986094>
40. Verney, S.P., Granholm, E., Marshall, S.P.: Pupillary responses on the visual backward masking task reflect general cognitive ability. *Int. J. Psychophysiol.* **52**, 23–36 (2004). <https://doi.org/10.1016/j.ijpsycho.2003.12.003>
41. Porter, G., Troscianko, T., Gilchrist, I.D.: Effort during visual search and counting: insights from pupillometry. *Q. J. Exp. Psychol.* **60**, 211–229 (2007). <https://doi.org/10.1080/17470210600673818>
42. Privitera, C.M., Renninger, L.W., Carney, T., Klein, S., Aguilar, M.: Pupil dilation during visual target detection. *J. Vis.* **10**, 1–14 (2010). <http://doi.org/10.1167/10.10.3>
43. Reiner, M., Gelfeld, T.M.: Estimating mental workload through event-related fluctuations of pupil area during a task in a virtual world. *Int. J. Psychophysiol.* **93**(1), 38–44 (2014)
44. Mathôt, S., Fabius, J., Van Heusden, E., Van der Stigchel, S.: Safe and sensible preprocessing and baseline correction of pupil-size data. *Behav. Res. Methods* **50**(1), 94–106 (2018). <https://doi.org/10.3758/s13428-017-1007-2>
45. Longo, L.: Subjective usability, mental workload assessments and their impact on objective human performance. In: IFIP Conference on Human-Computer Interaction, pp. 202–223. Springer, Cham (2017)
46. Longo, L.: Experienced mental workload, perception of usability, their interaction and impact on task performance. *PLoS ONE* **13**(8), e0199661 (2018)
47. Longo, L.: On the reliability, validity and sensitivity of three mental workload assessment techniques for the evaluation of instructional designs: a case study in a third-level course. In: International Conference of Computer Supported Education CSEDU (2) 2018, pp. 166–178 (2018)