



Task Demand Transition Rates of Change Effects on Mental Workload Measures Divergence

Enrique Muñoz-de-Escalona¹(✉), José Juan Cañas¹, and Jair van Nes²

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

² Faculty of Psychology and Neuroscience, University of Maastricht,
Maastricht, The Netherlands
vannes@student.maastrichtuniversity.nl

Abstract. Mental workload is a complex construct that may be indirectly inferred from physiological responses, as well as subjective and performance ratings. Since the three measures should reflect changes in task-load, one would expect convergence, yet divergence between the measures has been reported. A potential explanation could be related to the differential sensitivity of mental workload measures to rates of change in task-load transitions: some measures might be more sensitive to change than the absolute level of task demand. The present study aims to investigate whether this fact could explain certain divergences between mental workload measures. This was tested by manipulating task-load transitions and its rate of change over time during a monitoring experiment and by collecting data on physiological, subjective, and performance measures. The results showed higher pupil size and performance measure sensitivity to abrupt task-load increases: sensitivity to rates of change could partially explain mental workload dissociations and insensitivities between measures.

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

1 Introduction

The study of mental workload is of crucial importance in many fields such as Emergency Room healthcare or air traffic control (ATC), in which the lives of countless people are at stake and dependent on human performance [1, 2]. Human related accidents address an ongoing problem in social sciences and cognitive ergonomics: how can we minimize and avoid human error?

A high mental workload generally leads to poor performance [3, 4], and extreme cases of overload may result in errors which can in turn lead to fatal accidents. On the other hand, mental underload remains as undesirable as mental overload, likewise leading to poor performance and errors [5]. Automation, which does have its benefits,

is significantly and repeatedly associated with mental underload [6], and can cause a major loss of situation awareness, making it difficult for people to detect flaws and intervene adequately and timely. Therefore, optimizing levels of mental workload is vital to maintaining effective performance, avoiding both overload and underload during task performance. Furthermore, high levels of mental workload can have detrimental effects on people's psychosocial and physical health. High mental workload has been associated with high work-related fatigue, high-stress complaints, and/or burnout [7], as well as high scores on health complaint questionnaires [8]. Therefore, by using mental workload measures, we can learn its limits and dimensions and appropriately discover how to improve performance and minimize human error within organizations or within people's own personal work practices [4]. Task demand is dynamic in many fields such as ATC and aircraft pilots: workers may experience changes in mental workload as, for example, traffic load gets higher or a sudden unexpected storm appears when flying an aircraft, respectively. These changes in task-load may also be gradual or abrupt and affect individuals' mental workload, as well as the way its measures reflect these changes. To better understand and assess the construct of mental workload, the present study aims to investigate whether differential sensitivity to rates of change in task demand transitions would affect the convergence of mental workload measures. However, we will begin by defining mental workload as it in itself is a very poorly defined construct. Furthermore, we will define convergence and divergence phenomena between mental workload measures (associations, dissociations, and insensitivities), as well as the above-mentioned sensitivity to rates of change, to better understand the current mental workload assessment literature. The rest of the paper proceeds as follows. Section 2 outlines related work about general mental workload, focusing on defining and measuring mental workload and on introducing sensitivity to rates of change in task demand transitions. Section 3 describes the design and the methodology followed in conducting the experiment. Section 4 presents the obtained results, while Sect. 5 presents a discussion about our findings, as well as limitations and possible future work. Finally, Sect. 6 concludes the study, summarizing its key findings and implications on the extant body of knowledge.

2 Related Work

2.1 Defining Mental Workload

Mental workload is a complex construct without a clear consensus regarding a definition [9–11]. It has been defined as the product of the immediate demands of the environment and an individual's maximum mental capacity [12, 13], hence mental workload is a multi-factorial construct, which depends not only on demanded task resources but also on those available [14–16]. When the demands of the environment exceed a person's maximum mental capacity, mental overload occurs and performance deteriorates as a consequence of our limited capacity [17]. This is because when at the very limit of our human mental resources, we are unable to reallocate these resources in an adaptable way. On the contrary, when the environment demands too little, such as in work situations that are heavily automated, mental underload occurs and performance

similarly deteriorates. Why low environmental demands are detrimental is still poorly understood, however, some have suggested this may be due to shrinkages in our maximum mental capacity in response to environmental demand reduction [5, 18]; this in turn can influence several factors, including vigilance, workload, attention, and situation awareness [6]. For the purpose of our research, mental workload may be considered as the amount of mental effort involved in performing a task [3, 19]. In other words, the amount of mental resources in use during the performance of a task given the demands of the environment. The term task-load refers to environmental demands and it is used to manipulate the amount of experienced mental workload. In general, we say that measures reflecting mental workload are valid when they reflect a change in task-load (the demands of the immediate environment).

2.2 Measuring Mental Workload

The measurement of mental workload is one of the biggest challenges facing psychology and social sciences at present. There is a widespread need to assess cognitive work as, on the one hand, it is fundamental to the development of modern society and, on the other, it has been identified as one of the main causes of work-related accidents. Despite mental workload not being a directly measurable construct, it can be assessed with three types of individual “primary measures” reflecting mental workload [9, 11, 16, 20]: (a) physiological responses (Electroencephalography (other brain imaging techniques, heart-rate variability (HRV), pupil diameter, etc.); (b) perceived or subjective perception of mental workload (questionnaire or scale response); and (c) task performance (response speed and accuracy). According to Hancock (2017), if these three measures of mental workload mean to assess the same construct then they should demonstrate convergence. In other words, if task-load were to increase then we would expect the following associations between task-load changes and primary measures: (a) higher physiological activation responses, (b) higher perceived mental workload, and (c) a decrease in task performance. Thus, there should exist a convergence between these measures of mental workload, given the expected association of task-load with each respective measure. However, current research has demonstrated that this is not always the case: dissociations and insensitivities between mental workload measures have been repeatedly reported [16, 20, 21]. Insensitivities occur when a certain task has distinct levels of task-load, but measures of mental workload fail to reflect a change regardless of task-load levels. For example, when piloting an airplane, pilots may have to deal with dynamic changes within the immediate environment. Task-load may increase in line with the increasing demands of a complex situation, yet our measures reflect static levels of reported mental workload. For instance, task-load may increase due to air turbulence or even a failure in automation, yet physiological measures (such as pupil dilation or HRV) reflect no change. One must note, however, that it is possible for an insensitivity to occur for one measure of mental workload but not for the other. Given the current example, it may be that pilots’ physiological measures reflect no change in their actual mental workload, yet they can report an increase (or even decrease) in their perceived mental workload. Furthermore, dissociations occur when we have contradictory results: task-load increases but subjects report a lower perceived mental workload, whereas normally one would expect an increase in perceived mental

workload if task-load were to increase. Using our pilot example again, the situation at hand may increase task-load but pilots report (a) lower physiological activation responses (dissociation), (b) higher perceived mental workload (association), and (c) static levels in task performance (insensitivity). There are several factors that might affect the occurrence of dissociation and insensitivities between task-load and primary mental workload measures [20]. One possible explanation might be related to the timescale considered between measures. Muñoz-de-Escalona & Cañas (2018) identified that dissociations and insensitivities may appear at high mental workload peak experiences due to latency differences between measures: subjective measures showed lower levels of latency response than physiological response (pupil size) [16]. Despite this, there are also several other factors that might contribute to the emergence of divergence between mental workload measures, including sensitivity to rates of change [20].

2.3 Task Demand Transitions and Sensitivity to Rates of Change

Task demand transitions research has been very limited and largely focused on its effects in task performance and mental workload perceptions [22]. The current literature has revealed that a change in task-load levels affects mental workload, fatigue, and performance ratings [22–25]. However, the authors could not find any research focusing on how sensitivity to rates of change in task demand transitions affects the convergence and divergence of mental workload measures. Behavioral sciences has repeatedly shown that humans are more sensitive to change rather than the absolute level of a stimulus [20]. If we translate this into the study of mental workload convergence and divergence between measures, it may be possible that there exists a differential sensitivity to task-load rate of change in mental workload measures. Some measures might be more sensitive to change than to the absolute level of task demand, while others might be sensitive only to absolute levels of task demand. These differences would result in dissociations and insensitivities which would ultimately affect convergence between mental workload measures. This study aims to shed light on the effects of sensitivity to rates of change in task demand transitions on the convergence of mental workload measures. In the present study, we manipulated two independent variables: (1) task-load rate of change and (2) task-load change direction, whose effects were tested in a task-battery experiment in which participants were trained and instructed to perform to the best of their abilities. Participants performed the task-battery for 20 min, whilst data on the dependent variables, task performance, pupil diameter, and perceived mental workload were obtained. We hypothesized that there would be higher divergence (dissociations and/or insensitivities) between mental workload measures with abrupt rate of change conditions rather than linear ones.

3 Design and Methodology

3.1 Materials and Instruments

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 [25]. 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 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 ± 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.

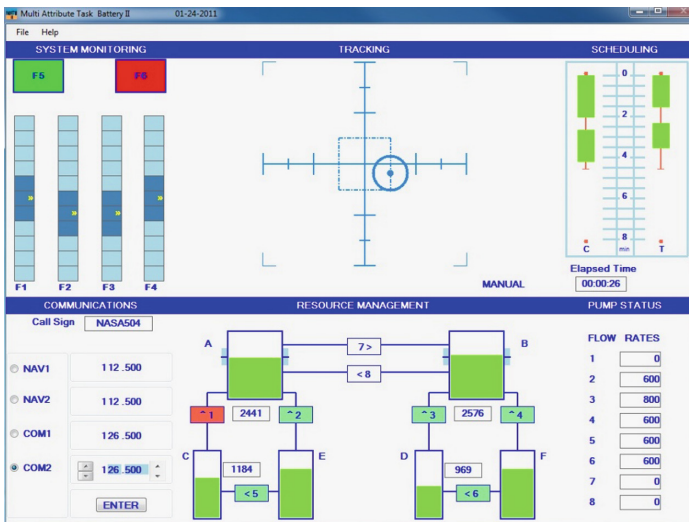


Fig. 1. MATB-II task display. Taken from <https://matb.larc.nasa.gov/>

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 [26].



Fig. 2. Tobii T120 Eyetracker system

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 experience on a scale ranging from 1 (no mental workload) to 5 (maximum mental workload), presented from left to right in ascending order of mental workload experienced. 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 [27, 28].

Participante:

Edad:

ISA SCALE

Responda según las siguientes opciones:

1 (Escasa de carga), 2 (Algo de carga), 3 (Carga moderada), 4 (Bastante carga), 5 (Máxima carga)

2.5 MIN	_____	1	2	3	4	5
5 MIN	_____	1	2	3	4	5
7.5 MIN	_____	1	2	3	4	5
10 MIN	_____	1	2	3	4	5
12.5 MIN	_____	1	2	3	4	5
15 MIN	_____	1	2	3	4	5
17.5 MIN	_____	1	2	3	4	5
20 MIN	_____	1	2	3	4	5

Fig. 3. Instantaneous self-assessment scale

3.2 Participants

Fifty-six psychology students from the University of Granada participated in the study. Participants’ ages ranged from 18 to 30, with an average of 22.7 and a standard deviation of 4. A total of 39 women and 17 men participated. It should be noted that there is a greater number of female participants due to the fact that psychology students at the University of Granada are mostly women. Recruitment was achieved through the dispersion of posters and flyers around the university, as well as an advertisement for the study on the university’s online platform for experiments (<http://experimentos.psiexpugr.es/>). The requirements for participation included (1) not being familiar with the MATB-II program, (2) Spanish 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. Participants’ participation was rewarded with two experimental vouchers for which they received extra credit.

3.3 Procedure

1. **Training stage:** training took place for no longer than 30 min. The objective of this stage was for participants 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 instructed to read the MATB-II instruction manual and inform the researcher once they had finished. 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 each MATB-II task separately and were first given a demonstration as to how to execute

the task and given time to perform the task themselves. 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 the participants 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 of three different rooms equipped for training with the MATB-II software, and no special attention to room conditions was needed.

2. **Data collection stage:** the data collection stage lasted approximately 20 min and involved participants completing 1 of the 4 randomly 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 and a half minutes when a scheduled alarm sounded. Prior to the start of the task-battery, the eye-tracker system was calibrated, and the participants were told to keep head and body movements to a minimum. During the data collection stage, standardizing room conditions was essential. 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:

In the present study we manipulated 2 independent variables:

- *Task-load rate of change:* this is the intensity in which task demand changes over time. We manipulated the task-load rate of change by modifying the number and the combination of active tasks that participants had to perform over time; this occurred during the data collection stage. We established 2 levels: (1) the variable rate of change and (2) the linear rate of change. Possible task combinations are illustrated in Table 1.

Table 1. Possible task combination in MATB-II software

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

- *Task-load change direction*: this is the direction in which task-load changes over time. Since task demand transitions can occur in two directions, we manipulated this variable on 2 levels: (1) increasing task-load change and (2) decreasing task-load change.

As a result of the manipulation of these two variables, we obtained 4 experimental conditions (see Fig. 4), namely:

1. **Condition: Increasing task-load with a variable rate of change.** Task-load increased every 5 min, but with a variable rate of change. Participants had to perform 1, 2, 4, and 5 sets of task combinations: there is an initial rate of change from task combination 1 to 2, then there is an abrupt increase in the rate of change from task combination 2 to 4 and then a decrease in the rate of change from task combination 4 to 5.
2. **Condition: Decreasing task-load with a variable rate of change.** Participants performed the same set of tasks from condition (1) but in descending order, resulting in 5, 4, 2, and 1 task combinations.
3. **Condition: Increasing task-load with a linear rate of change.** Task-load increased every 5 min with a linear rate of change. Participants had to perform task combinations 1, 2, 3, and 5.
4. **Condition: Decreasing task-load with a linear rate of change.** Participants performed the same set of tasks from condition (3) but in descending order, resulting in task combinations 5, 3, 2, and 1.

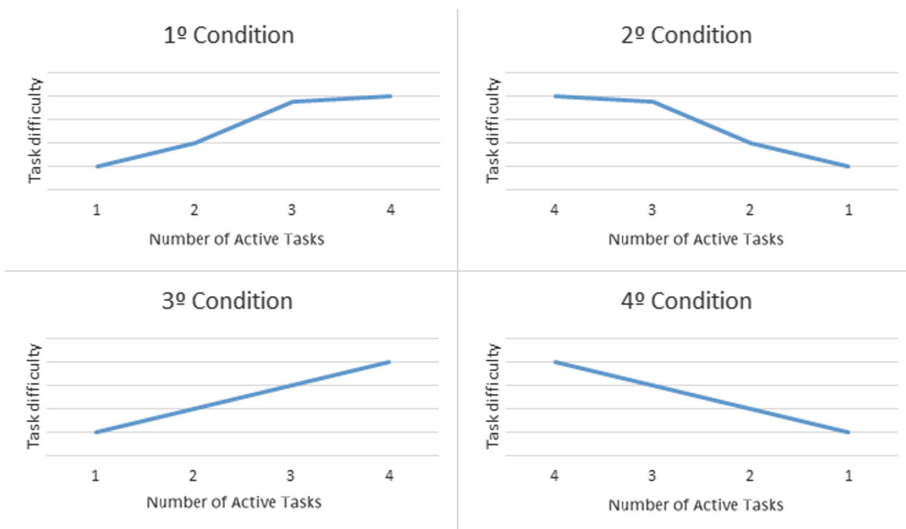


Fig. 4. Task-load rate of change evolution for experimental conditions.

Note that the difference between the increasing variable and linear rate of change conditions demonstrates that the former condition involves a sharp increase in task-demands from task combination 2 to 4. The reason for this lies with the elimination of an already practiced task (TRACKING) and the addition of 2 new non-practiced tasks (COMM & RESMAN), whereas the latter conditions involve a linear increase in task-demands from task combination 2 to 3 since only a single new task is added (COMM) and vice-versa regarding a decreasing variable and a linear rate of change conditions.

Dependent Variables:

Performance. MATB-II provides us with many indicators of participants' performance: e.g. 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 SYSMON performance indicator, as it is the only task present during all 4 of the task-load levels in the 4 experimental conditions, allowing us to compare participants' performances between conditions. The SYSMON performance indicator will be considered as the number of correct responses divided into the number of possible responses. The result is a number between 1 (best possible performance) and 0 (worse possible performance).

Pupil Size. Mental workload can be reflected 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 effectively reflects mental workload [29–37] and minimize intrusiveness. While our eye-tracking system allows continuous sampling rate recording at 120 Hz, we set a total of 8 intervals lasting 2.5 min each in order to obtain 2 measures per task-load level. 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 [38] 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 allowing us to reduce noise in our 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 Taskload 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-min intervals throughout the 20 min of the experimental stage, obtaining a total of 8 subjective mental workload ratings (2 measures per task demand level).

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 would also allow the ISA scale to be synchronized with the performance and pupil size measures.

4 Results

We used three one-way, within-subjects ANOVA to analyze the obtained results, one for each mental workload measurement. First, the analyses of participants' performance showed that our task-load level manipulation was successful. The ANOVA analyses identified a very significant main task-load level effect $F(3,156) = 74.34$, $MSe = .005$, $p < .01$, which reflects that participants' performances decreased as task demand increased. The main effect of the experimental condition was found to be significant $F(3,52) = 3.24$, $MSe = .02$, $p < .01$. Moreover, an interaction effect of the task-load level x experimental condition was also found to be significant $F(9,156) = 6.14$, $MSe = .005$, $p < .01$. This demonstrates that performance variations evolve differently throughout the different task-load levels depending on the considered experimental condition: there is a higher decrease in performance in the variable rate of change conditions compared to the linear rate of change conditions from task-load levels 2 to 3 (see Fig. 5).

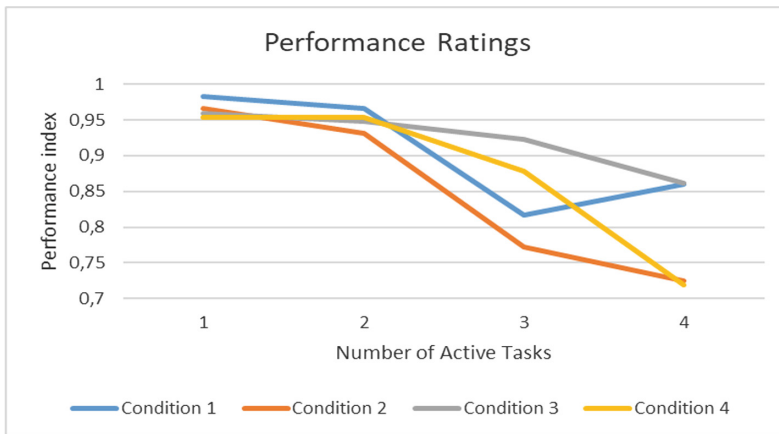


Fig. 5. Participants' performance during task development.

Regarding subjective perceptions, a linear increase in participants' subjective mental workload ratings occurs between task-load levels in every experimental condition: in the ANOVA, the main effect of task-load levels turned out to be very significant indeed, $F(3,156) = 358.9$; $MSe = .240$, $p < .01$; whereas the main effect of the experimental condition $F(3,52) = 1.09$; $MSe = .84$, $p > .05$ and the interaction of

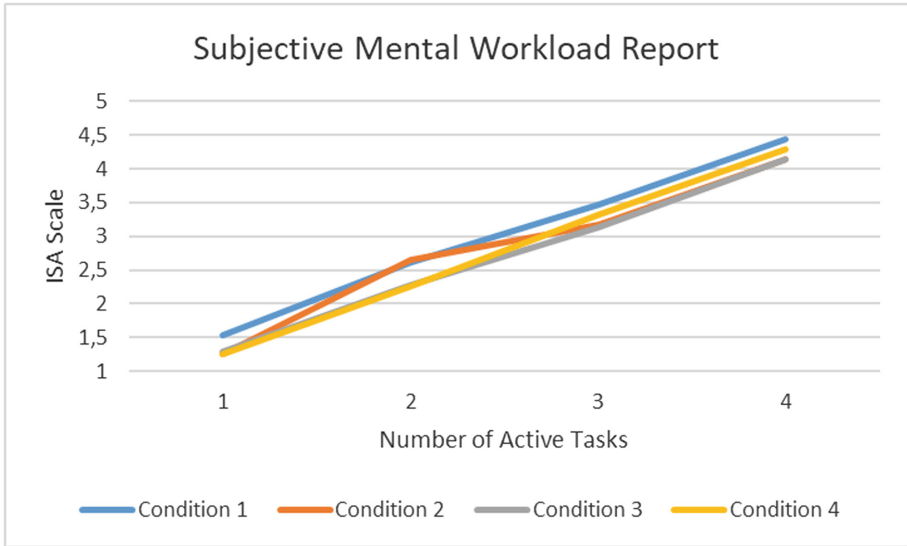


Fig. 6. Participants' subjective mental workload ratings during task development.

the task-load level \times experimental condition $F(9,156) = .72$; $MSe = .240$, $p > .05$ were not significant (see Fig. 6).

For pupil size, our physiological measurement, we also found a very significant main effect of task-load level $F(3,156) = 70.94$, $p < .01$ which supported participants' pupil size increasing (higher activation) as task demand rose. Despite the main effect of the experimental condition not being significant $F(3,52) = 1.03$, $MSe = .00$, $p > .05$, a significant interaction effect of the task-load level \times experimental condition was found $F(9,156) = 4.93$; $MSe = .008$, $p < .01$. This implies that pupil size variation evolves differently through task-load levels depending on the considered experimental condition: in decreasing task-load conditions (2 & 4), pupil size increases linearly with higher task demand, regardless of the task-load rate of change; whereas in increasing task-load conditions (1 & 3), pupil size starts at a higher state of dilation (task-load level 1), then decreases in the following task-load level (task-load level 2). From task-load level 2 to 3, there is a higher dilation in the variable rate of change condition (1) regarding the linear condition (3). Finally, from task-load level 3 to 4, pupil dilation decreases in the variable rate of change condition (1), whereas it continues increasing in the linear rate of change condition (3) (see Fig. 7).

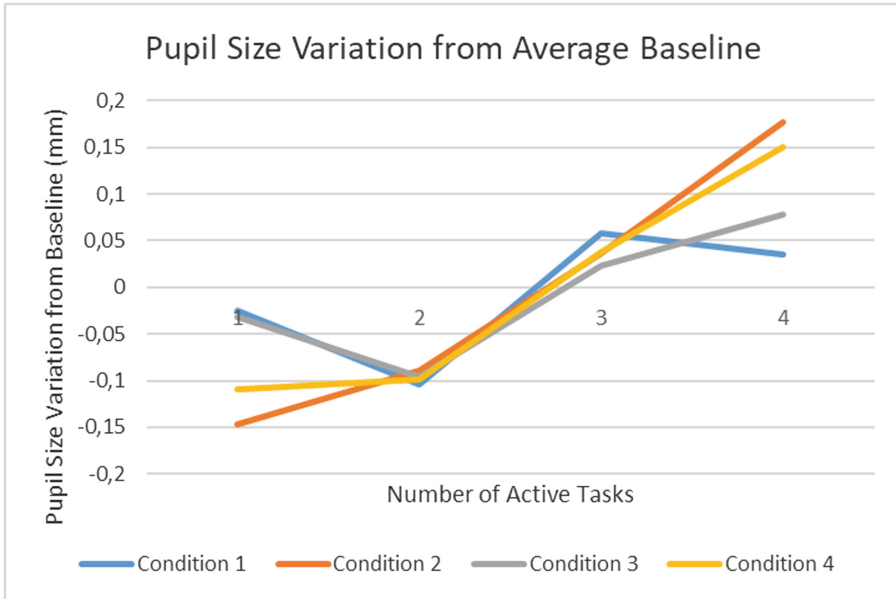


Fig. 7. Participants' pupil size variation ratings from the average baseline during task development.

5 Discussion

To synthesize our data, we can see how subjective ratings increased linearly as task-load rose in every experimental condition, whereas the performance and physiological measures reacted differently depending on the experimental condition in question. Performance linearly decreased as task-load increased in both (increasing and decreasing) linear rate of change conditions (3 & 4), as well as in the decreasing variable rate of change condition (2). Conversely, in the increasing variable rate of change condition (1), performance decreased until task-load level 3 and then improved again from task-load levels 3 to 4. Pupil size increased linearly in both decreasing task-load conditions (2 & 4) as task demand increased; while in both increasing task-load conditions there was a decrease in pupil size from task-load levels 1 to 2, followed by an increase from task-load levels 2 to 3. Finally, from task-load levels 3 to 4, pupil size continued increasing in the linear rate of change condition but decreased in its variable rate of change counterpart.

Hence, according to our data and in terms of association, dissociation, and insensitivities:

- Regarding *subjective response*, we found associations between task-load and subjective perceptions in every experimental condition: there is a direct mapping between mental workload subjective ratings and task demand. We can observe how subjective ratings have not been influenced either by task-load rate of change or task-load change direction.

- Regarding *performance response*, our results presented associations or dissociations depending on the considered experimental condition: we found associations between task-load and performance responses in both (increasing and decreasing) the linear rate of change conditions and in the decreasing variable rate of change condition; performance became more linearly impaired as task-load increased. However, in the increasing variable rate of change condition, we found associations between task demand and performance responses until task-load level 3, but an improvement in performance was observed from task-load levels 3 to 4 (dissociation) which contradicted the task-load increase.
- Concerning our *physiological response*, we also found associations or dissociations depending on the considered experimental condition: our results discovered associations between pupil size variations and task demand in both decreasing task-load conditions: pupil size increased linearly with higher task demand. However, considering increasing task-load conditions we found dissociations from task-load levels 1 to 2 (pupil size decreases) and associations from task-load levels 2 to 3 (pupil size increases) in both variable and linear rate of change conditions. To contrast, from levels 3 to 4, we again found a dissociation in the variable rate of change condition (pupil size decreases), but an association with the linear rate of change condition (pupil size increases).

Therefore, this data partially confirmed our hypothesis that there exists a higher divergence between mental workload measures with variable rate of change conditions: we found performance dissociation only in the increasing task-load with a variable rate of change condition (condition 1). Furthermore, we found physiological dissociations in both variable rate of change conditions from task-levels 1 to 2, but from task-load levels 3 to 4 we found a dissociation only in the increasing variable rate of change condition. Hence, taking into account our results, we could say that there exists a higher divergence between mental workload measures with a variable rate of change condition, particularly in increasing task-load variable rate of change conditions. These results could be explained under the explanation provided by Hancock (2017, page 12) in which he claims that ...[one of the more well-established principles that we do have in the behavioural sciences is that human frequently prove more sensitive to change rather than the absolute level of a stimulus array]... [20]. In other words, some mental workload reflections could be more sensitive to change rather than the absolute level of task demand, whereas others might be more sensitive to the absolute levels of task demand. Therefore, by analyzing our results, subjective perception ratings appeared to be less sensitive to task demand rate of change: they linearly increased in every experimental condition showing no statistical differences among them. Nevertheless, performance measures appeared to be more sensitive to an abrupt increase in task-load level, as the worst performance peak in the increasing variable rate of change condition was achieved in the abrupt transition from task-load levels 2 to 3, even though task demand was higher in task-load level 4. Moreover, physiological measures also appeared to be more sensitive to abrupt changes in task demand only in the increasing task-load condition. Similarly to what happened with performance measures, a higher pupil size peak was achieved in the abrupt transition from task-load levels 2 to 3, despite task-load being higher in task-load level 4. As pupil dilation reflects activation

(among other factors which were controlled), when there is an abrupt increase in task demand, we seem to overreact in order to prepare ourselves to face environmental threats but, in line with resources theories, due to the fact that mental resources are limited and can be depleted, when an abrupt increase is followed by a soft increase in task-load, our organism detects that there is no need to continue activating and it deactivates in order to save resources. On the other hand, higher pupil size in task-load level 1 for both increasing task-load conditions (compared to task-load level 2 in the same condition, and compared to task-load level 1 in the decreasing task-load conditions) could be explained by the fact that participants' activation were higher at the beginning of the experimental session because of the natural nervousness experienced by participants. While in both the decrement task-load conditions, this nervousness activation is added to the activation produced by task-demand level 4, as is reflected in our data: pupil dilation for decreasing task-load conditions were higher in task-load level 4 than for increasing task-load conditions. These findings should be viewed in light of some study limitations. Although we were able to overcome other studies' limitations, such as the examination of a single direction task demand transition [39], because of the high number of possible combinations, it was not suitable to analyse other interesting experimental conditions (as in for example, low-high-low/high-low-high transitions or changing the intervals in which the abrupt change in task-load demand occurs: beginning, middle, and end of the scenario). Moreover, we think it would be highly interesting to introduce other physiological measures in this study, such as EEG and/or HRV. We must bear in mind that divergences have been found not only between the three primary mental workload measures (performance, physiological, and subjective), but within different indexes of each primary indicator. Lastly, this study has been conducted with students under simulated conditions and we think that it would be appropriate to validate these findings under real conditions in order to improve ecological validity.

Further research is needed to untangle mental workload divergence between measures. Future research could address the aforementioned methodological limitations. For example, it would be interesting to analyse how sensitivity to the rate of change effects varies depending on when the abrupt change takes place or how low-high-low/high-low-high affects mental workload measures' convergence.

6 Conclusions

Mental workload is a complex construct which can be measured by its three primary measures: performance, physiological, and subjective. Despite expecting to find convergence between them as they reflect the same construct, dissociations and insensitivities have been repeatedly reported in the literature. A potential explanation for these divergences could be related to the differential sensitivity of mental workload measures to task-load transitions rate of change: some measures might be more sensitive to change than the absolute level of task demand, while others might be more sensitive to absolute levels of task demand. These differences would result in dissociations and insensitivities, which would ultimately affect convergence between mental workload measures. Our results suggest that dissociations in performance and physiological pupil

size measures may appear after an abrupt change takes place, albeit mostly during increasing task-load conditions. However, subjective ratings may not be affected by the task-load rate of change but by the absolute level of task demand. In other words, our results partially confirmed our hypothesis, as we found higher divergence (dissociations and/or insensitivities) between mental workload measures with abrupt rates of change but only during the increasing task-load condition. An important implication of this finding is that we should give more weight to one of the mental workload reflections depending on environmental rate of change demands. In other words, if task-load transitions are linear, then we could rely on every primary mental workload measure; but when there is an abrupt task-demand transition from low to high mental workload, we may prefer to rely on subjective ratings rather than physiological or performance measures, as the subsequent decrement in physiological activation (saving resources) would not necessarily mean a decrement in an operator experienced mental workload.

References

1. Pape, A.M., Wiegmann, D.A., Shappell, S.A.: Air traffic control (ATC) related accidents and incidents: A human factors analysis (2001)
2. Reuters: What We Know About the Deadly Aeroflot Superjet Crash Landing, 6th May 2019. <https://www.themoscowtimes.com/2019/05/06/what-we-know-about-the-deadly-aeroflot-superjet-crash-landing-a65495>
3. Byrne, A.J., Sellen, A.J., Jones, J.G.: Errors on anaesthetic record charts as a measure of anaesthetic performance during simulated critical incidents. *Br. J. Anaesth.* **80**(1), 58–62 (1998)
4. Byrne, A.: Measurement of mental workload in clinical medicine: a review study. *Anesth. Pain Med.* **1**(2), 90 (2011). <https://doi.org/10.5812/kowsar.22287523.2045>
5. 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.ub.unimaas.nl/10.1080/14639220210123789>
6. Endsley, M.R.: From here to autonomy: lessons learned from human–automation research. *Hum. Factors* **59**(1), 5–27 (2017). <https://doi-org.ezproxy.ub.unimaas.nl/10.1177/2F0018720816681350>
7. Zoer, I., Ruitenburg, M.M., Botje, D., Frings-Dresen, M.H.W., Sluiter, J.K.: The associations between psychosocial workload and mental health complaints in different age groups. *Ergonomics* **54**(10), 943–952 (2011). <https://doi.org/10.1080/00140139.2011.606920>
8. Kawada, T., Ooya, M.: Workload and health complaints in overtime workers: a survey. *Arch. Med. Res.* **36**(5), 594–597 (2005). <https://doi-org.ezproxy.ub.unimaas.nl/10.1016/j.arcmed.2005.03.048>
9. Cain, B.: A review of the mental workload literature. Defence Research and Development Toronto (Canada) (2007)
10. Meshkati, N., Hancock, P.A. (eds.): *Human Mental Workload*, vol. 52. Elsevier, Amsterdam (1988)
11. Moray, N. (ed.): *Mental Workload: Its Theory and Measurement*, vol. 8. Springer, New York (1979). <https://doi.org/10.1007/978-1-4757-0884-4>

12. Kantowitz, B.H.: Attention and mental workload. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 44, no. 21, pp. 3–456, July 2000. <https://doi-org.ezproxy.ub.unimaas.nl/10.1177%2F154193120004402121>
13. Wickens, C.D.: Multiple resources and mental workload. *Hum. Factors* **50**(3), 449–455 (2008). <https://doi.org/10.1518/001872008X288394>
14. Wickens, C.D.: Multiple resources and performance prediction. *Theor. Issues Ergon. Sci.* **3**(2), 159–177 (2002). <https://doi.org/10.1518/001872008x288394>. 2008 50:449
15. Muñoz-de-Escalona, E., Canas, J.: Online measuring of available resources. In: H-Workload 2017: The First International Symposium on Human Mental Workload, Dublin Institute of Technology, Dublin, Ireland, 28–30 June (2017). <https://doi.org/10.21427/d7dk96>
16. Muñoz-de-Escalona, E., Cañas, J.J.: Latency differences between mental workload measures in detecting workload changes. In: Longo, L., 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
17. Durantin, G., Gagnon, J.F., Tremblay, S., Dehais, F.: Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behav. Brain Res.* **259**, 16–23 (2014). <https://doi.org/10.1016/j.bbr.2013.10.042>
18. Young, M.S., Stanton, N.A.: Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Hum. Factors* **44**(3), 365–375 (2002). <https://doi.org/10.1518/0018720024497709>
19. Byrne, A.J., et al.: Novel method of measuring the mental workload of anaesthetists during clinical practice. *Br. J. Anaesth.* **105**(6), 767–771. <https://doi.org/10.1093/bja/aep268>
20. Hancock, P.A.: Whither workload? Mapping a path for its future development. In: Longo, L., Leva, M. Chiara (eds.) H-WORKLOAD 2017. CCIS, vol. 726, pp. 3–17. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61061-0_1
21. Yeh, Y.Y., Wickens, C.D.: Dissociation of performance and subjective measures of workload. *Hum. Factors* **30**(1), 111–120 (1988). <https://doi-org.ezproxy.ub.unimaas.nl/10.1177%2F001872088803000110>
22. Edwards, T., Martin, L., Bienert, N., Mercer, J.: The relationship between workload and performance in air traffic control: exploring the influence of levels of automation and variation in task demand. In: Longo, L., Leva, M. (eds.) H-WORKLOAD 2017. CCIS, vol. 726, pp. 120–139. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61061-0_8
23. Helton, W.S., Shaw, T., Warm, J.S., Matthews, G., Hancock, P.: Effects of warned and unwarned demand transitions on vigilance performance and stress. *Anxiety Stress Coping* **21**, 173–184 (2008)
24. Cox-Fuenzalida, L.E.: Effect of workload history on task performance. *Hum. Factors* **49**, 277–291 (2007)
25. Santiago-Espada, Y., Myer, R.R., Latorella, K.A., Comstock Jr., J.R.: The multi-attribute task battery II (MATB-II) software for human performance and workload research: a user’s guide (2011)
26. Lee, J., Ahn, J.H.: Attention to banner ads and their effectiveness: an eye-tracking approach. *Int. J. Electron. Commer.* **17**(1), 119–137 (2012). <https://doi.org/10.2753/JEC1086-4415170105>
27. 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)
28. 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)

29. Matthews, G., Middleton, W., Gilmartin, B., Bullimore, M.A.: Pupillary diameter and cognitive and cognitive load. *J. Psychophysiol.* **5**, 265–271 (1991)
30. Backs, R.W., 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)
31. Hyönä, J., Tommola, 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>
32. Granholm, E., Asarnow, R.F., Sarkin, A.J., Dykes, K.L.: Pupillary responses index cognitive resource limitations. *Psychophysiology* **33**, 457–461 (1996). <https://doi.org/10.1111/j.1469-8986.1996.tb01071.x>
33. 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>
34. 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>
35. 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>
36. Priviter, C.M., Renninger, L.W., Carney, T., Klein, S., Aguilar, M.: Pupil dilation during visual target detection. *J. Vision* **10**, 1–14 (2010). <https://doi.org/10.1167/10.10.3>
37. 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)
38. 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>
39. Morgan, J.F., Hancock, P.A.: The effect of prior task loading on mental workload: an example of hysteresis in driving. *Hum. Factors* **53**(1), 75–86 (2011)