Theoretical and Methodological Considerations in the Comparison of Performance and Physiological Measures of Mental Workload

Julien Cegarra[1] and Aline Chevalier[2]

[1] CLLE-LTC Laboratoire Travail et Cognition (UMR CNRS 5263), Centre Universitaire Jean-François Champollion, Place de Verdun, 81012 Albi cedex 9, France
[2] Cognitive Processes and Interactive Behaviours Laboratory, University of Paris X-Nanterre, 200 avenue de la République, 92001 Nanterre cedex, France
Julien.Cegarra@univ-jfc.fr, Aline.Chevalier@u-paris10.fr

Abstract. Mental workload is a central concept to a range of disciplines (including cognitive ergonomics) and has been the subject of various debates. As a result, a number of different techniques have been used by researchers. In this study, we focus on the comparison of performance and physiological measures of mental workload. We describe a puzzle-solving experiment in which we combined multiple measures of mental workload. The results highlight a number of issues. We stress several theoretical and methodological questions that may help to reach a clearer understanding of the nature of the mental workload concept.

Keywords: mental workload; dissociation; dual-task paradigm; eye dilation.

1 Introduction

Mental workload is a central concept to a range of disciplines. For example, in cognitive ergonomics there is a need to regulate the human operator's mental workload in order to prevent both underload and overload, whilst maintaining an adequate level of performance. At the same time, the diverse theoretical foundations of mental workload are still under debate. As a result, a number of different techniques have been used by researchers. Methods for measuring mental workload are usually classified using antitheses, such as direct *versus* indirect, objective *versus* subjective or analytical *versus* empirical. One of the most intuitive classifications focuses on the means for assessing mental workload, taking into consideration performance measures (e.g., dual-task paradigm), subjective measures (e.g., rating scales) and physiological measures (e.g., pupil dilation). Relatively few researchers have attempted to compare methods from different categories (but see, for instance, [7]). Here, we suggest some theoretical and methodological considerations in the comparison of performance and physiological measures.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 264–268, 2007.

2 Performance and Physiological Measures of Mental Workload

2.1 Performance Measures

Many studies indicate a strong link between poor performance in the task and the mental workload experienced by the participant. However, this approach is not always reliable as the relationship between mental workload and performance also depends on task demands, human strategies and individual differences (e.g., [2]). Therefore, different researchers suggested to measure the workload from an additional task. The most frequently used approach to additional task measurement is the dual-task paradigm: Participants are asked to complete a task and at various times are interrupted by a signal (e.g., a tone), which they have to respond as rapidly as possible. It is assumed that the amount of time necessary to respond to the signal reflects the amount of mental workload allocated to the main task. Although some researchers have challenged the assumptions made by this method (e.g., [4]), many others have gone on to use it successfully in a range of tasks, including design activities or information-searching (e.g., [3]).

2.2 Physiological Measures

Physiological measures (such as the measurement of variations in heartbeat, pupil dilation, blink or respiration rate) are based on the assertion that physiological variables reflect changes in cognitive functioning. In this way, the pupil dilation technique is considered to be very sensitive in many tasks and is, therefore, seen to be one of the most relevant physiological measure [1]. Moreover, pupil dilation is considered precise enough to assess subtle variations in workload. However, researchers noted the lack of selectivity of physiological measures (e.g., luminosity is also known to be a factor determining pupil size). In this study, we are comparing this technique to a performance measure, which is more selective but with a lower bandwidth.

3 Method

We designed an experiment based on the Sudoku puzzle (see Fig 1). The aim is to fill a 9 by 9 grid with numerical digits so that every row, every column and every 3 by 3 box contains the digits from 1 to 9. Each participant had to solve 2 puzzles with 2 levels of complexity. During puzzle-solving, we recorded the participant's pupil dilation using the SMI iViewX head-mounted eye-tracking system with a 50 Hz sampling rate. The ambient light was controlled during the experiment, which took place in a noise-attenuated room. As the participant performed the task, we recorded reaction times to standard tones played through the computer's speakers at 10 to 20 second intervals. The intrusiveness of the signal range was controlled in a pilot study.

Fig. 1. The Sudoku puzzle: The participant has to fill the grid, selecting the numerical digits so that every row, every column and every 3 by 3 box contains the digits from 1 to 9

4 Results

To precisely compare the two techniques, we focus the analysis on the easiest of the two puzzles on one participant. Figure 2 displays two curves along a temporal axis: (1) the reaction times on a scale from 0 to 500 msecs, displayed as a curve (rather than discrete plots) in order to facilitate comparison; (2) pupil diameter on a scale from 3 to 4.5 mm, as well as a smoothed curve of these values.

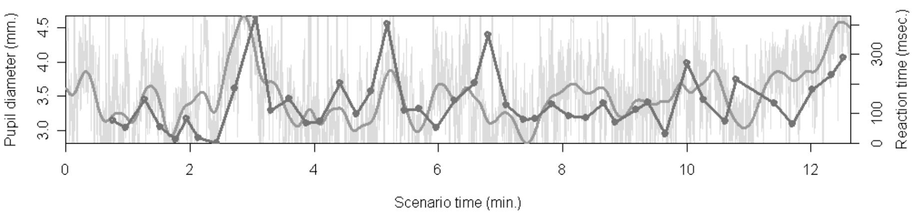

Fig. 2. Curves of the reaction times (msec.) and pupil dilation measures (mm.) for the easiest puzzle

A comparison between the two curves cannot be made on exact values as the unit measurements differ for each scale. However, the shape of the curves provides relevant information. On closer examination of the two curves, differences in shape become apparent:

- The peaks do not appear exactly at the same moment. The peaks in the pupil dilation curve sometimes precede (e.g., at about 3 minutes on the x-axis), seem to just follow (e.g., at about 5 minutes) or seem unrelated (e.g., at about 7 minutes) to a corresponding peak in the reaction time curve.

- The physiological measure captures more changes. Even while looking at the smoothed curve of pupil dilation, the curve of the reaction time often badly approximates its shape. So the sensitivity of the curve is indeed dependent on the sampling of the measure.

5 Discussion

These results illustrate the benefits of comparing physiological and performance measures of mental workload.

From a methodological point of view, it is almost impossible to discuss the changes in mental workload for the reaction times without using another measure. For example, care must be observed when claiming that an increase in workload is followed by a decrease (or vice versa) because the shape of the curve is uncertain between the known values. Another consideration results from the diagnosticity of the eye dilation measurement technique. As Wickens and Hollands [8] stated, physiological measures have to be considered within the context of separate pools of resources. The apparent differences in the peaks between the eye dilation measurement technique and the dual-task paradigm may result from the fact than the eye dilation may be more diagnostic of the load being considered. As noted by Wilson and O'Donnell [9], an increased load does not always result in increased overall activation. The more diagnostic measure would detect changes in the load, whereas the general measure of load would not. This would also mean that even if the changes in the two graphs seem to imply a close correlation, one has to be careful interpreting the results as the changes do not necessarily have a common cause.

Moreover, from a theoretical point of view, questions arise concerning the validity of the reaction times to measure workload. It is necessary to distinguish between delays in reaction times which are related to workload and delays related to conflicting attentional demands or psychological refractory period (i.e. the signal is appearing when a response to the main task is underway). A more detailed comparison of physiological measures and the dual-task paradigm (with a larger participant sample), as suggested by Isreal et al. [6], might provide more details about the theoretical foundations of the dual-task paradigm. This may take the form of a comparison between the dual-task paradigm to the irrelevant-probe technique. This technique involves recording of physiological measures (such as ERP) in which, unlike the dual-task paradigm, the participant has to ignore the probes. Several authors have noted that attention not focused elsewhere is attracted by the probes even if the participant does not directly pay attention to the probe [5].

6 Conclusion

These results stress that such analysis of performance and physiological measures are crucial, especially since many studies measure mental workload with an additional task. For example, mental workload in car driving is evaluated whilst receiving and holding conversations on a mobile phone (i.e. measuring gaps in the telephone conversation).

Therefore, more studies are necessary to precisely identify the theoretical and methodological limitations of performance and physiological measures of mental workload.

References

1. Beatty, J.: Task evoked pupillary responses, processing load, and the structure of processing resources. Psych. Bull. 91, 276–292 (1982)
2. Cegarra, J., Hoc, J.M.: Cognitive Styles as an Explanation of Experts' Individual Differences: A Case Study in Computer-Assisted Troubleshooting Diagnosis. Int. J. of Hum.-Comp. Stud. 64, 123–136 (2006)
3. Chevalier, A., Kicka, M.: Web designers and web users: Influence of the ergonomic quality of the web site on the information search task. Int. J. of Hum.-Comp. Stud. 64, 1031–1048 (2006)
4. Fisk, A.D., Derrick, W.L., Schneider, W.: A Methodological Assessment and Evaluation of Dual-Task Paradigms. Cur. Psych. Res. & Rev. 5, 315–327 (1987)
5. Kramer, A.F., Trejo, L.J., Humphrey, D.: Assessment of mental workload with task-irrelevant auditory probes. Biol. Psych. 40, 83–100 (1995)
6. Isreal, J.B., Chesney, G.L., Wickens, C.D., Donchin, E.: P300 and Tracking Difficulty: Evidence for Multiple Resources in Dual-Task Performance. Psychophysiology 17, 259–273 (1980)
7. Miyake, S.: Multivariate workload evaluation combining physiological and subjective measures. Int. J. of Psychophysiology 40, 233–238 (2001)
8. Wickens, C.D., Hollands, J.: Engineering psychology and human performance. Longmans, London, UK (2000)
9. Wilson, G.F., O'Donnell, R.D.: Measurement of operator workload with the neuropsychological workload test battery. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 63–100. Elsevier, Amsterdam (1988)