Estimating Sub-categories of Cognitive Load: An Eye-tracking Study

Sonali Aatrai
Advanced Technology & Development Centre
Indian Institute of Technology Kharagpur
West Bengal, India

Sparsh Kumar Jha
Department of Electrical Engineering
Indian Institute of Technology Kharagpur
West Bengal, India

Rajlakshmi Guha
Rekhi Centre of Excellence for the Science of Happiness
Indian Institute of Technology Kharagpur
West Bengal, India

Abstract

This study explores eye-markers for sub-categories of cognitive load. Experiments were conducted on 63 participants using Image Sliding Puzzle (ISP). NASA-TLX was administered post task completion as a measure of cognitive load. Total scanning duration, total fixation duration, fixation count and total saccadic duration were found to be significant, which is consistent with pre-existing literature. Next, we investigated whether sub-categories of cognitive load (mental demand, temporal demand, perceived performance, effort, and frustration) can be distinguished by characteristic eye-metrics. Our findings reveal signature eye-markers for specific sub-categories of cognitive load. Further, we explored the link between perceived performance and actual performance and established that mean fixation duration, peak velocity, mean saccadic duration, and skewness in saccadic velocity were significant markers for both objective and subjective markers of performance. To our knowledge this is the first study to compare the task-evoked eye measures for sub-categories of cognitive load.

Keywords: Eye-tracking; Cognitive load; Problem-Solving; Performance;

Introduction

Cognitive load is a construct that represents the amount of processing resources required by a given task. Cognitive load is essential to research (Plass, Moreno, & Brünken, 2010) because it reveals how the human brain processes information and handles mental effort throughout tasks. A large body of resources have been engaged in studying cognitive load and its effect on task performance. Sweller (1988) labelled cognitive load into three types based on the processes involved. Intrinsic load, is directly related to the task. The second category, extraneous load, is caused by the format and presentation of information. Whereas, understanding and manipulating the content results in germane load.

Complexity of the information being processed and the amount of resources required to process it can both have an effect on cognitive load (Wang, Yang, Liu, Cao, & Ma, 2014). A low extraneous load increases cognitive capacity for intrinsic and germane load, which aids in the completion of the current activity. There is an inverse link between cognitive load and performance. Performance declines as cognitive burden rises (Sweller, 1988). When the cognitive load exceeds a person's ability to absorb information and accomplish a task, it results in lower performance, difficulties with information retention, and diminished learning outcomes. In contrast, when cognitive load is minimal, an individual is able to absorb information more efficiently, resulting in higher performance, better information retention, and enhanced learning outcomes. Individuals' processing capacity and performance on a complex task vary systematically.

Subjective and objective markers refer to different methods of measuring a phenomenon or concept. Subjective markers are self-reported measures, such as questionnaires, interviews, or rating scales, where individuals provide their own perceptions and opinions about a particular concept, such as their experience of cognitive load. On the other hand, objective markers are measures that are independent of an individual's perceptions or opinions, and are based on observable and quantifiable data. In the context of cognitive load, completion of the task is widely used objective marker along with steps taken and reaction times (Wang et al., 2014). Both subjective and objective markers have their own strengths and limitations. Subjective markers are typically more accessible and less invasive, but may be affected by individual biases and memory limitations. Objective markers, on the other hand, provide more reliable and quantifiable data, but may not always capture the subjective experience of the individual.

There are several markers of cognitive load that are commonly used in research studies, including physiological measures (heart rate, skin conductance, and electroencephalogram), behavioural measures (reaction time, error rate, number of steps), self-report measures (subjective ratings of workload) and eye movement analysis (Fraser, Ayres, & Sweller, 2015).

The NASA-TLX (NASA Task Load Index) is widely used as a validated subjective tool (questionnaire) to measure and assess the cognitive load of a task or system (Hart, 2006). It

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

assesses mental, physical, and emotional demands on the user by considering six sub-aspects: mental demand, physical demand, temporal demand, own perception of performance, effort, and frustration. The NASA-TLX involves asking the individual to rate the mental, physical and temporal demands of a task, as well as their frustration and performance level (Zagermann, Pfeil, & Reiterer, 2016). The scores from these ratings are combined to produce a single overall score, which can be used to compare the cognitive load experienced across different tasks and systems. It was developed by NASA to evaluate the subjective workload of astronauts during space missions, but it has since been widely used in various fields including human factors, ergonomics, and psychology to assess workload in various tasks and environments. However, it has some limitations. Because it relies on participant self-report, it cannot provide real-time information about cognitive load. Additionally, it may not be relevant for evaluating unconscious or automated processes. Combining an objective metric with the subjective marker of NASA TLX has been the preferred assessment tool for cognitive load in most studies (Zheng et al., 2012; Ikuma, Harvey, Taylor, & Handal, 2014; Luro & Sundstedt, 2019).

Eye tracking provides real-time data about cognition by measuring the visual behaviour through eye movements and gaze patterns during a task or activity. Eye movements, can be both voluntary (e.g. fixations and saccades) and involuntary (e.g. pupil dilation), and have been found to be related to cognitive processes (Nam, 2020). Increasingly, eye movements are employed to explore how visual perception and visual search influence cognitive processes. Eye parameters can be used to quantify a person's cognitive load when solving puzzles or other types of problems. Previous studies have shown that when cognitive load increases, eye movements become less stable and more variable, indicating a decrease in the efficiency of visual processing. Additionally, increased cognitive load is associated with longer fixation duration, indicating that more time and effort is being used to process information. Eye tracking can also be used to study the relationship between cognitive load and visual attention (Wang et al., 2014). For example, research has shown that when cognitive load is high, people tend to focus on the most important or relevant information, and ignore less important or irrelevant information. This can be observed by measuring the number of fixations and saccades on different parts of a display.

The NASA-TLX questionnaire and eye-tracking are useful tools for analyzing cognitive load. NASA-TLX is a comprehensive measure of workload but its limitations in real-time analysis can be overcome by eye-tracking (Zagermann et al., 2016), which can assess cognitive load as it occurs in real-life situations. Both methods have their own strengths and limitations, and a combination of techniques may be necessary to fully understand cognitive load in different scenarios. There is a pressing need for a non-invasive measure of individuals' cognitive load, to avoid overloading users and therefore non-invasive eye movement analysis is gaining ground in the study of cognitive load research. This study bridges the gap between NASA-TLX and eye-tracking by examining the unique eye markers associated with various sub-components of cognitive load, including mental demand, physical demand, temporal demand, perceived performance, effort, and frustration. This integration provides a more comprehensive understanding of cognitive load and its impact on human performance.

Motivation

Performance includes two measures: completion time and error rates. Adding a third measure, cognitive load provides a more complete picture of performance by taking into account the mental demand, physical demand, temporal demand, own perception of performance, effort, and frustration required to complete a task (Zagermann et al., 2016).

In this study of cognitive load, a combination of both subjective and objective markers is used to provide a more comprehensive understanding of the phenomenon, as they complement each other and provide different perspectives on the same experience. The goal of this study is to bridge the gap between cognitive load and eye-tracking by examining the unique eye markers associated with various sub-components of cognitive load, including mental demand, physical demand, temporal demand, perceived performance, effort, and frustration.

Objective

1. To explore if there are eye-metrics that are significant markers of the overall cognitive load.

2. To find if there are eye-metrics that are significant markers of the parameters of the cognitive load in NASA-TLX.

3. To investigate if there is a link between subjective marker and objective marker of performance, i.e., perceived and actual performance, respectively.

Methods

Ethical Clearance & Participants

The design of this experiment received clearance from the ethical committee of institute (No. IIT/SRIC/DR/2019) to proceed with the experiment. For this study, 63 university students (31 females and 32 males; age group: 21–36 years with a mean of 27.08 and standard deviation of 3.94) were selected after screening. Participants whose weighted gaze was less than 80% were excluded during screening. The participants in this study had a normal or corrected-to-normal vision and reported having no history of neurological or psychological problems. Before participating, everyone provided informed consent.

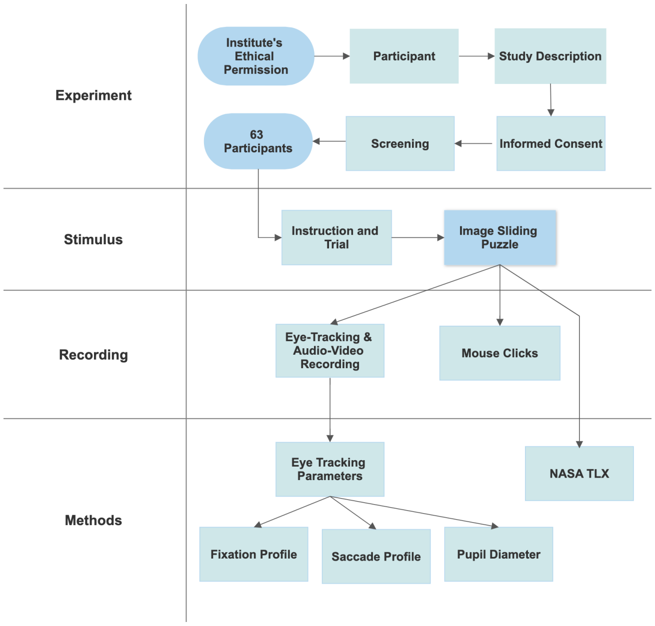

Figure 1: Flow of study

Experimental Setup & Method

Before beginning the data-collecting process, the eyes of each participant were calibrated with the use of an eye-tracking device (Model: Tobii Pro X3-120) that was mounted below an HP 24f display (24 inches, 60 Hz) with a screen resolution of 1920 x 1080 pixels. The participants were seated 60 cm from the display in a closed-door, noise-free room. Each participant was given an Image Sliding Puzzle (n x m), which is an arrangement problem which divides the reference image in n x m similar sized pieces whose positions are then jumbled and one of the pieces are taken out. The goal of the problem is to reorganise the bits such that they reach the reference picture while adhering to the fundamental instructions, which are as follows: Each move consists of exchanging the empty position with its surrounding pieces, i.e. the empty position may only be switched with the piece on its top, bottom, left, and right if it is present. To get participants accustomed to the platform, they were given a simple 2x2 initial problem in which they were taught the rules and how the Image Sliding Puzzle task worked. The main problem (3x3 Image Sliding Puzzle (ISP), with an optimal number of moves of 15) was presented to the participants thereafter. Performance based responses and total time taken in the tasks were recorded. Tobii X3-120 was used to capture eye parameters during task performance. Simultaneously, audio and video were recorded during the task for further checks. Further, after completion of the task, the participants were administered NASA TLX questionnaire.

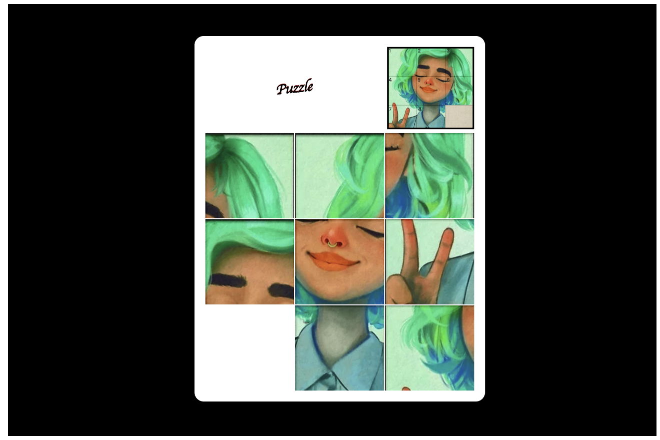

Figure 2: Image Sliding Puzzle (ISP)

Feature Extraction & Analysis

Our pre-processing methodology was divided into four stages (Kiefer, Giannopoulos, Raubal, & Duchowski, 2017). The initial stage was to remove any missing data points, such as eye blinks and out-of-focus gaze locations. The pixel coordinates of two successive gaze samples were then utilised to calculate the Euclidean distance between them. Third, the time between samples was calculated by subtracting the timestamp of the previous sample from the timestamp of the present sample. Finally, the gaze velocities were calculated by dividing the distance travelled by the eyes by the time difference between samples. The data was analysed using custom Python programs. The data were analysed using Pearson Correlation and Analysis of Variance (ANOVA) unless otherwise mentioned.

From the eye tracker and mouse data, various eye and perfromance parameters were obtained. These parameters are:

1. Total Fixation Duration (TFD): It is the aggregate of all of the fixation periods that occurred during the session.

2. Fixation Count (FC): The total number of fixations that occurred during the session

3. Mean Fixation Duration (MFD): The average of all fixation periods during the task.

4. Time to First Fixation (TFF): Time taken before the first fixation occurs.

5. First Fixation Duration (FFD): The time duration of the first fixation.

6. Peak Velocity (PV): It is the maximum gaze velocity attained throughout the task.

7. Mean Saccadic Duration (MSD): The average of all saccadic periods during the task.

8. Mean saccadic amplitude (MSA): The euclidean distance that separates two successive fixation locations is referred to as saccadic amplitude. MSA represents the average of all saccadic amplitudes.

9. Mean saccadic Velocity (MSV): The average of all saccadic velocities during the task.

10. Skewness of Saccadic Durations (SSD): Using Equation 1, the skewness of saccadic durations is computed.

$$Skewness = \frac{\sum_i^N (Y_i - \mu)^3}{(N-1) * \sigma^3} \quad (1)$$

Where, Y_i = random variable, μ = mean of the distribution, σ = standard deviation, N = number of variables in the distribution,

11. Skewness of Saccadic Velocity (SSV): Similarly, using Equation 1, the SSV was calculated from saccadic velocities.

12. Total Saccadic Duration (TSD): It is the aggregate of all of the saccadic periods that occurred during the session.

13. Total Scanning Duration (TScD): Entire duration of the task is TScD.

14. Mean pupil diameter (MPD): Average of pupil width during entire session is called MPD. Usually, it attains a size of 2 to 4 mm.

15. Skewness in pupil diameter (SPD): The skewness of pupil diameter is calculated using Equation 1.

16. Reaction Time (RT): It is the time taken by participant to make the first move.

17. Clicks (CLK): It is the total number of clicks in the task. It indicates the number of steps.

Results & Discussion

Performance is a function of correct responses, number of moves and time taken. Cognitive load is an important performance indicator because it provides a more complete picture of performance by accounting for the mental demand, physical demand, temporal demand, own perception of performance, effort, and frustration required to complete a task. In ISP, 32 participants completed the task, while 31 participants were unable to do so. Pearson bivariate correlation method was applied to examine the relationship between the cognitive load parameters obtained by NASA TLX and eye-metrics along with other markers such as reaction time and clicks. Table 1 lists the eye-markers that were found to be significantly correlated with various cognitive load parameters.

Previous studies (Krejtz, Duchowski, Niedzielska, Biele, & Krejtz, 2018; Wang et al., 2014; Klingner, 2010) shows that total scanning duration, total fixation duration, fixation count and total saccadic duration are signature markers of cognitive load and have positive correlation. Our findings on overall cognitive load from NASA-TLX are consistent with these previous findings. It shows that total scanning duration (Pearson correlation: 0.278, Significance: 0.038), total fixation duration (Pearson correlation: 0.269, Significance: 0.045), fixation count (Pearson correlation: 0.284, Significance: 0.034) and total saccadic duration (Pearson correlation: 0.264, Significance: 0.048) have significant positive correlations with the cognitive load.

Table 1: Correlation between NASA-TLX Parameters and Eye and Performance metrics

Type	Features	Pearson Correlation	Significance (2-tailed)
Mental Demand	TScD	0.311	0.02
	CLK	0.341	0.01
	TFD	0.308	0.021
	FC	0.309	0.02
	TSD	0.293	0.028
	MPD	0.326	0.014
Temporal Demand	RT	0.343	0.01
	FFD	-0.272	0.043
	PV	0.305	0.022
	SSD	0.272	0.042
	SSV	0.271	0.043
Perceived Performance	FFD	0.294	0.028
	MFD	-0.318	0.017
	PV	0.344	0.01
	MSD	0.325	0.015
	SSD	-0.317	0.017
	SSV	0.311	0.02
	SPD	0.319	0.017
Effort	FFD	-0.264	0.049
Frustation	SSV	0.32	0.016
	MSD	-0.263	0.05
	MSV	0.269	0.045
Overall Cognitive Load	TScD	0.278	0.038
	TFD	0.269	0.045
	FC	0.284	0.034
	TSD	0.264	0.048

We further examine to see if there exist signature markers for various sub-components of cognitive load. Total scanning duration, click count, mean pupil diameter, total fixation duration, fixation count, total saccadic duration, and total fixation duration were found to be significantly correlated with mental demand. All of these parameters had a positive correlation, indicating that higher values of these parameters are associated with higher levels of mental demand. We observed that reaction time, first fixation duration, peak velocity, skewness in saccadic duration and skewness in saccadic velocity are significantly correlated with temporal demand. All these parameters had positive correlation, indicating that more temporal demand is associated with higher values of these parameters except first fixation duration. It was seen that the first fixation duration is lesser for higher temporal demand. Likewise, first fixation duration, mean fixation duration, peak velocity, mean saccadic duration, skewness in saccadic duration, skewness in saccadic velocity and skewness in pupil diameter are significantly correlated with perceived performance. All these parameters had positive correlation, indicating that poor perceived performance is associated with higher values of these parameters except mean fixation duration and skewness in saccadic duration. It was seen that the mean fixation duration and skewness in saccadic duration is more for better perception of performance. Effort showed a negative correlation with first fixation duration, indicating, more effort is associated with lesser first fixation duration. Skewness in saccadic velocity and mean saccadic velocity is positively correlated with frustration, whereas mean saccadic duration is negatively correlated. It must also be noted that no parameters were found significantly correlated with physical demand.

Next, we investigated if there is a link between subjective marker and objective marker of performance, i.e., perceived and actual performance, respectively. We divided the group in two subgroups, i.e., good performers (those who completed the task, N = 30) and poor performers (those who did not complete the task, N = 33). The results in Table 2 show a significant difference (F = 4.326, P = 0.042) in perceived performance among good performers (Mean = 3.966, SD = 2.4125), and poor performers (Mean = 5.222, SD = 2.0817), confirming that actual performance (completion of a task) is in sync with perceived performance (here lower number indicate better perceived performance on a scale of 1-10).

Table 2: Significant Eye-parameters based on performance (*At 95% CL, Rest all at 99% CL)

Features	Mean	Std. Dev.	F Value	Significance
Perceived Performance	3.966 5.222	2.4125 2.0817	4.326	0.042*
MFD	273.19 229.29	46.1037 48.0435	12.172	0.001
PV	14.982 19.876	3.3532 8.1303	8.894	0.004
MSD	32.82 34.808	1.8173 2.5072	11.662	0.001
SSV	1.258 1.691	0.3438 0.6465	9.995	0.003

In addition, using the results, we were able to derive the mean fixation duration, peak velocity, mean saccadic duration, and skewness in saccadic velocity as signature markers (with a confidence level of more than 99 percent) of performance in this task. It's interesting to note that these ocular markers had a substantial correlation with people's ratings of their performance in the NASA-TLX test. Poor performers had significantly higher peak velocity, mean saccadic velocity, and skewness in saccadic velocity, whereas good performers had significantly longer mean fixation duration. Peak velocity, mean saccadic velocity, and skewness in saccadic velocity were significantly higher in poor performers.

From our finding, we propose signature markers of different parameters for cognitive load, such as, mental demand, temporal demand, perceived performance, effort, frustration. We also validated the pre-existing literature, establishing a link between overall cognitive load and the following eye-markers - total scanning duration, total fixation duration, fixation count and total saccadic duration. In addition to this, we found a correlation between the subjective marker of perceived performance and the objective marker of actual performance. As a result, we also validated the signature eye-markers for perceived performance.

Conclusion

The findings suggest that eye-tracking technology can be a valuable tool for monitoring cognitive load during visual problem solving tasks. Eye movements are highly correlated with an individual's subjective experience of cognitive load and their objective performance on a task. The study reveals that eye metrics of total scanning duration, total fixation duration, fixation count, and total saccadic duration are significant markers of cognitive load. Signature markers have also been identified for mental demand, temporal demand, perceived performance, effort, and frustration as subcategories of cognitive load. The study also found a link between perceived performance and actual performance, suggesting that eye-tracking measures can provide valuable insights into the individual's perceived experience of cognitive load. Mean fixation duration, peak velocity, mean saccadic duration, and skewness in saccadic velocity were found to be significant markers for both objective and subjective markers of task performance. The results of this study can be a forerunner of future studies in eye movement analysis and cognitive load. The implications of this research also highlight the importance of developing methods for measuring cognitive load that does not disrupt user workflow or performance, particularly in critical tasks such as driving, machine operation, maritime bridge operations, and cybersecurity operations. The pursuit of this objective has the potential to enable more effective workload monitoring, which could improve performance and safety outcomes in these domains. Additionally, this research offers opportunities for the development of machine learning games. However, further research is needed to validate the signature markers of other cognitive load parameters, as well as explore the potential limitations and sources of variability in eye-tracking data.

References

Fraser, K. L., Ayres, P., & Sweller, J. (2015). Cognitive load theory for the design of medical simulations. *Simulation in Healthcare*, *10*(5), 295–307.

Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, pp. 904–908).

Ikuma, L. H., Harvey, C., Taylor, C. F., & Handal, C. (2014). A guide for assessing control room operator performance using speed and accuracy, perceived workload, situation awareness, and eye tracking. *Journal of loss prevention in the process industries*, *32*, 454–465.

Kiefer, P., Giannopoulos, I., Raubal, M., & Duchowski, A. (2017). Eye tracking for spatial research: Cognition, computation, challenges. *Spatial Cognition & Computation*, *17*(1-2), 1–19.

Klingner, J. (2010). *Measuring cognitive load during visual tasks by combining pupillometry and eye tracking*. Stanford University.

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PloS one*, *13*(9), e0203629.

Luro, F. L., & Sundstedt, V. (2019). A comparative study of eye tracking and hand controller for aiming tasks in virtual reality. In *Proceedings of the 11th acm symposium on eye tracking research & applications* (pp. 1–9).

Nam, C. S. (2020). *Neuroergonomics*. Springer.

Plass, J. L., Moreno, R., & Brünken, R. (2010). Cognitive load theory.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*(2), 257–285.

Wang, Q., Yang, S., Liu, M., Cao, Z., & Ma, Q. (2014). An eye-tracking study of website complexity from cognitive load perspective. *Decision support systems*, *62*, 1–10.

Zagermann, J., Pfeil, U., & Reiterer, H. (2016). Measuring cognitive load using eye tracking technology in visual computing. In *Proceedings of the sixth workshop on beyond time and errors on novel evaluation methods for visualization* (pp. 78–85).

Zheng, B., Jiang, X., Tien, G., Meneghetti, A., Panton, O. N. M., & Atkins, M. S. (2012). Workload assessment of surgeons: correlation between nasa tlx and blinks. *Surgical endoscopy*, *26*, 2746–2750.