

Measuring Cognitive Load

Antoine Richermoz, Vera Filipovich

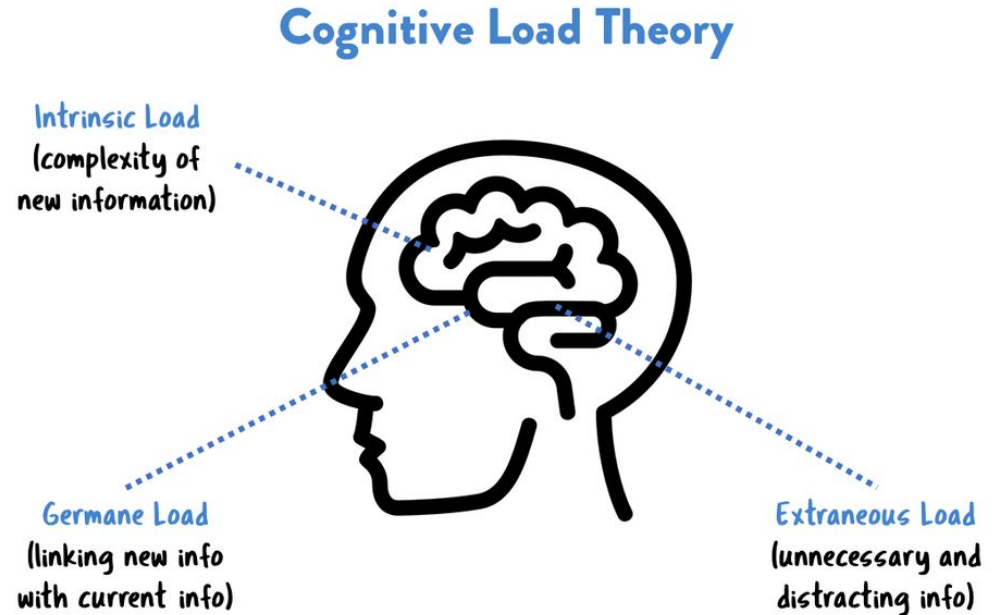


Summary

- Definition
- Objectives
- Methods
- Advantages and drawbacks

What is cognitive load ?

Cognitive (work)load is an overall amount of mental resources dedicated to performing a task.



Why is it important to evaluate?

- 1) Some usability tests may fail in real-world conditions due to stress, fatigue, lack of sleep, or many other cognitive factors.



<https://usabilitygeek.com/remote-usability-testing-best-practices/>

Why is it important to evaluate?

2) Some methods can give an idea of which areas of the app/website or which situations are particularly difficult for users



<https://business-antidote.com/optimisation-technique-contenu-referencer-site/>

Three types of methods

- Physical measurements
- Subjective post-event tests
- Indirect measurements

Physical measurements

Physical measurements are based on the factors that are independent of the user's will such as:

- eye movement
- body temperature
- heart rate

Physical measurements: Eye movements

Kosch, T., Hassib, M., Woźniak, P. W., Buschek, D., & Alt, F. (2018). *Your Eyes Tell*

Shows how cognitive workload leads to increased gaze differences of smooth pursuit eye movements during the presence of cognitive workload.

Task: N-back recall

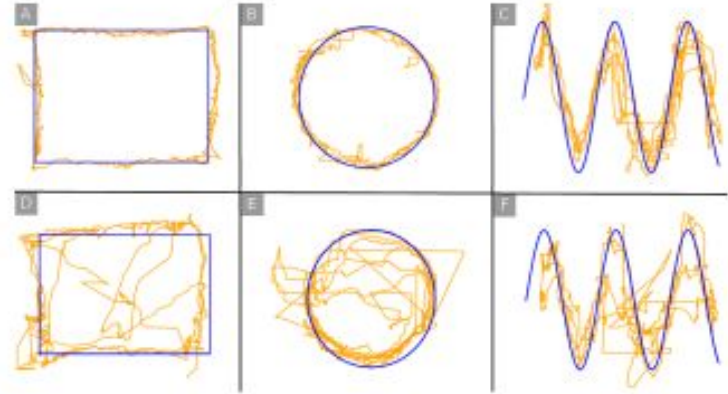


Figure 1. Smooth pursuit recordings under different levels of cognitive workload for three trajectories. The blue line shows the displayed trajectory, while the orange line visualizes the gaze path. (A), (B), and (C) show the gaze path during low cognitive workload phases, while (D), (E), and (F) shows the gaze path during perception of high workload.

Physical measurements: Thermal imaging

Stemberger, J., Allison, R. S., & Schnell, T. (2010). *Thermal Imaging as a Way to Classify Cognitive Workload*

Proposed a system based on thermal infrared imaging of the face, head pose estimation, measurement of the temperature variation across regions of the face.

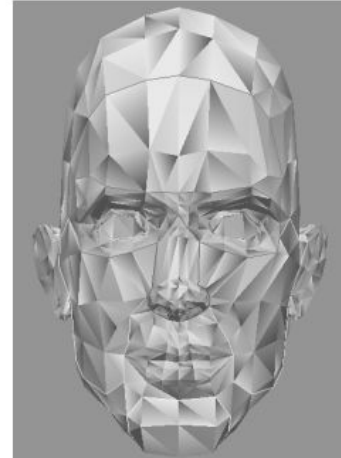


Figure 1. Head model used for all participants



Figure 2. Approximate ROI as seen on a thermal image

Physical measurements: pros & cons

Pros:

- Objective measurements
- Direct access during a specific task (ex .eye measurements for monitoring task)

Cons:

- Often performed in laboratory conditions
- Highly depended on physiological processes

Subjective post-event tests

Ask users to fill out forms to measure their subjective cognitive workload

May consist of:

- Scales
- MCQ
- ...

1. How satisfied are you with the current features of the products in this category?

- Dissatisfied
- Not very satisfied
- Moderately satisfied
- Very satisfied
- Extremely satisfied

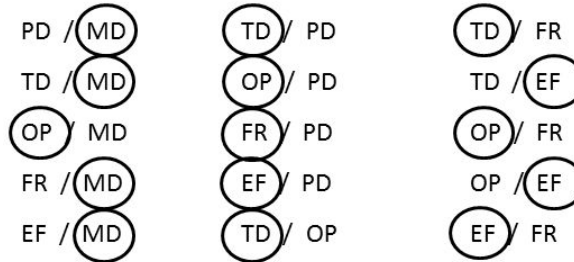
2. Rate the factors that affect your buying decision for this category.

Price of the products	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7
Promotions on the products	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7
Store preference	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7
Brand	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7
Quality	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7
Earlier usage	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7
New trials	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7

Subjective post-event tests: NASA TLX

Consist of two parts:

1. Six scales
2. Comparison of the importance of those 6 scales



NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date
Mental Demand How mentally demanding was the task?		
Very Low Very High		
Physical Demand How physically demanding was the task?		
Very Low Very High		
Temporal Demand How hurried or rushed was the pace of the task?		
Very Low Very High		
Performance How successful were you in accomplishing what you were asked to do?		
Perfect Failure		
Effort How hard did you have to work to accomplish your level of performance?		
Very Low Very High		
Frustration How insecure, discouraged, irritated, stressed, and annoyed were you?		
Very Low Very High		

Subjective post-event tests: pros & cons

Pros:

- Measures the perceived cognitive workload
- Can be used in any situation
- Easy to use (official NASA TLX app)

Cons:

- Measurement is done after the task
- Is subjective

Indirect measurements

Indirect measurements include a secondary physical or cognitive task to perform by the participant.

- Sternberg Memory Test
- Tapping test

Indirect measurements: Sternberg Test

Was designed to measure how quickly people can search for and retrieve information from short-term memory.

Participant has to memorize a small set of number. While performing the task, a number appear on the screen at fixed intervals. If the prob number were in the set to memorize the response should be “yes”, otherwise the participant should say “no”.

The speed of response (saying yes/no) varies with the cognitive load.

Indirect measurements: Tapping Test

The subject is asked to tap with a constant rhythm. At the same time the participant is performing a task on a web site.

Areas where the rhythm of tapping is slowed down or stopped are subjects of examination.

Indirect measurements: pros & cons

Pros:

- Provide a clear indicator of high load situation/areas

Cons:

- Impossible to know how the second task overlaps with the primary one
- (Sternberg Memory Task) The prime task is interrupted when performing the secondary task

Sources

Casner, S.M., & Gore, B.F. (2010). Measuring and Evaluating Workload: A Primer.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load : Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, 100(1), 223-234.
<https://doi.org/10.1037/0022-0663.100.1.223>

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index) : Results of Empirical and Theoretical Research. *Advances in Psychology*, 139-183. [https://doi.org/10.1016/s0166-4115\(08\)62386-9](https://doi.org/10.1016/s0166-4115(08)62386-9)

Kosch, T., Hassib, M., Woźniak, P. W., Buschek, D., & Alt, F. (2018). Your Eyes Tell. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. Published. <https://doi.org/10.1145/3173574.3174010>

Stemberger, J., Allison, R. S., & Schnell, T. (2010). Thermal Imaging as a Way to Classify Cognitive Workload. 2010 Canadian Conference on Computer and Robot Vision. Published. <https://doi.org/10.1109/crv.2010.37>

Tracy, Janet & Albers, Michael. (2006). Measuring Cognitive Load to Test the Usability of Web Sites.