

# Towards Dynamic and AI-Based Multimodal Cognitive-Load Assessment with Low-Hardware Setup for Programming Learning

Pavel Kostarev<sup>ID</sup> Alexander Bartel<sup>ID</sup>

Neu-Ulm University of Applied Sciences, Germany

How to cite: Kostarev, P.; Bartel, A. (2026). Towards Dynamic and AI-Based Multimodal Cognitive-Load Assessment with Low-Hardware Setup for Programming Learning. In: 12th International Conference on Higher Education Advances (HEAd'26). Valencia, 15-18 June 2026. <https://doi.org/10.4995/HEAd26.2026.21660>

## Abstract

*In this paper, we address the issue of accessibility in personalized learning for undergraduate students in programming education. We have designed an adaptive learning framework that uses continuous data collection through eye tracking and mouse movements to measure overall cognitive load (CL) in short intervals. The aim is to provide students with constructive, continuous feedback in the form of brief hints generated by a Large Language Model (LLM). Our paper streamlines the methodology and makes the concept of adaptive learning approachable by designing a lightweight model for CL assessment via pupil dilation, blink rate and mouse click patterns. This reduces deployment costs, since the model can be run on any PC with a standard webcam. It also ensures privacy and empowers researchers with affordable tools for collecting longitudinal evidence for long-term studies.*

**Keywords:** Cognitive load estimation; Human-computer interaction (HCI) in education; Learning Analytics; Adaptive Learning; Eye-tracking; Mouse-movement analytics.

## 1. Introduction

In recent years the concept of adaptive learning has become more popular (Strielkowski et al., 2025). Personalization of studying materials, tasks, quizzes and learning curriculum facilitates the learning process, increasing the engagement and gains from learning materials (Wang et al., 2024). Educational content should no longer be just consumed by a student but rather adjusted in real-time to the current knowledge level, learning preferences, learner's abilities and environment (Martin et al., 2020). In the light of evolving research directions of generative and multimodal Artificial Intelligence (AI), the implementation of such systems in a practical setup became much easier. To personalize and adapt the learning materials displayed for students, we need to assess the problems which a student encounters when studying from the materials prepared by a teacher in a traditional (presentation-slides) setup. With the future target of

building an adaptive learning platform, in this paper we focus on the first paradigm in designing an adaptive platform, which involves detection and interpretation of difficulties in understanding learning materials indicated by higher CL, primarily. While conventional adaptive learning platforms are designed only considering direct interactions with the user interface (learning materials, subjective evaluation and cognitive tests) and further judgements based on these interactions, we would like to focus on user metadata, which are currently less understood and less employed in the real educational setup, but still carry valuable cognitive and psychological insights on perception of the learning content. Aligning with the existing research gaps in learning analytics, where psychological impacts and predictive modeling for learning content are understudied, our hypothesis is that dynamic assessment of user metadata, such as eye and mouse movements can deliver valuable insights for detecting parts of learning material, which were poorly understood and require to be either reformulated or further explained (Kabudi et al., 2021; Martin et al., 2020; Shemshack & Spector, 2020; Wang et al., 2024).

The research question that we pose in this work can be formulated as follows: *Can we design a framework that uses mouse activity and eye tracking data via a webcam to estimate total CL in real time, producing personalized hints facilitating the learning process for students using an LLM?*

## **2. Background and Related Work**

### **2.1. Cognitive load overview**

Cognitive load theory (CLT) provides useful insights into how our brains process information and how difficult learning materials are for working memory to process at once (de Jong, 2010). CLT includes three types of cognitive load: intrinsic (refers to the mental effort required by the natural complexity of the material.), extraneous (extra CL caused by instructional design components used for presenting the learning materials, such as a layout or a user interface) and germane (an effort to link new information into existing knowledge in memory, creating a virtual a “mind map”) (Sweller et al., 1998). In practical setups, it is difficult to distinguish between the three types of CL during measurement. Tools for the differentiated measurement of CL are still in the early stages of development and require a larger sample size for validation. Therefore, total CL is often used instead (Klepsch et al., 2017; Leppink, 2017).

We can measure the total CL either subjectively (by self-assessment) or objectively (behavioral and physiological methods).

### *2.1.1. Subjective methods*

Among the subjective methods, there are single-item mental effort scales (9-point Likert scales) developed by Fred Paas (Duran et al., 2022; Paas et al., 1994), which remains still very popular as well as multidimensional workload questionnaires such as the Subjective Workload Assessment Technique („The Subjective Workload Assessment Technique“, 1988) and the NASA Task Load Index („Development of NASA-TLX (Task Load Index)“, 1988), which extend the traditional Linkert scales with weighted score across several categories providing more detailed insights on the causes of higher CL.

While subjective methods cannot deliver accurate results due to noise and bias, physiological and behavioral methods were in the focus of cognitive studies for a long time.

### *2.1.2. Behavioral methods*

These methods measure the student's observable actions while they are working on the task. Examples include eye tracking, mouse activity (frequency and trajectory), cursor tracking, and changes in body posture (Asteriadis et al., 2009).

### *2.1.3. Physiological methods*

Electroencephalography (EEG), Pupillary Response, Blink Rate, Eye fixation duration, Galvanic Skin Response (GSR) and Heart Rate Variability (HRV) are often used for physiological measurement of CL. Comparison studies often emphasize that pupillometry dynamics are the most reliable physiological signals for assessing CL (Alshanskaia et al., 2024; Ayres et al., 2021).

There are several examples of employing both physiological and behavioral measurement in E-Learning frameworks. For instance, Herbig et al. was using Tobii Pro Glasses 2 eye-tracker and Empatica E4 wristband for tracking HRV and GSR together with Kinect v2 for body posture estimation to assess the cognitive load while watching instructional videos or solving quizzes. The results indicate that combination of several modalities, such as eye-tracking data with HRV and GSR for prediction of CL deliver the highest prediction accuracy. However, the authors also emphasize that body postures did not make any significant contribution to prediction of CL and rather performed poorly, while the eye-tracking data had the highest feature importance over GSR and HRV. In addition, the framework just assesses the CL without any further interaction with a student (providing useful hints to help with material marked as “poorly understood”). It also requires a range of additional hardware components, making it difficult to use this framework for every student at home (Herbig et al., 2020).

LLMs are very actively employed for the concept of adaptive tutoring and identifying gaps in knowledge as for instance in *PELICAN* project (*PELICAN*, 2025), where LLM generates diagnostic questions or *SocratiQ* project (Jabbour et al., 2025), which provides explanations on

demand. However, there are almost no existing tools or only few experimental projects combining learning hints from LLM together with multimodal input assessing the CL.

### 3. Methods

#### 3.1. System architecture

We developed a flask-based web application in Python for a browser-based learning environment for first-semester bachelor's students studying Python. The course is conducted in German. The application uses a webcam to track eye movements and record mouse movements simultaneously, to assess students' working and cumulative cognitive load while working with individual presentation slides. Our framework is based on experiments demonstrating that modern webcam-based trackers can achieve a level of accuracy comparable to that of laboratory-grade eye trackers, such as the EyeLink 1000 (Kaduk et al., 2024). This makes them suitable for the detailed study of workload at home. All system components, including calibration, tracking, key feature extraction and cognitive load assessment, operate in real time.

#### 3.2. Eye tracking with a webcam: calibration, validation and filtering

The webcam detects and tracks both pupils within each frame and converts their positions into normalized screen coordinates (pixels). Before starting the calibration phase, we make sure that the face is well visible in the camera, the distance to the screen is within the allowed range and that light conditions are well enough to continue. Next, to ensure the high data quality, we perform a multi-point calibration (11 points) comparable to those used in laboratory systems and recommended webcam tracking. We calibrate each point for 3 head positions (left, straight, right) to ensure the stability for further eye-tracking for any slight changes in the posture. Finally, the validation phase estimates the quality of the calibration by using the angular error at the validation points. Sessions with excessive error are flagged and asked to be repeated.

Raw gaze samples are then filtered using a Kalman filter. This is primarily necessary to suppress camera noise; head shake and transient spikes. Blinks are detected by characteristic patterns in the pupil signal, such as disappearance and reappearance of the pupil and significant changes in pupil confidence. Fixations and saccades are extracted from the smoothed gaze stream using velocity- and dispersion-based thresholds, in accordance with established eye movement analysis practice. For each fixation, we record the start and end time, its duration and centroid coordinates. Saccades are characterized by start time and their amplitude (Grabinger et al., 2024; Zagermann et al., 2016).

We compute the pupil diameter using PuRe method (Santini et al., 2018) per frame by averaging left and right pupil size continuously for every slide. Pupil baselines are estimated during an initial rest interval (directly after calibration step, during validation phase) and continuously

updated via a sliding window. For every slide we derive the following signals: a.) Baseline-corrected mean pupil diameter (task-evoked dilation), b.) Pupil variance and a pupil velocity (mean absolute frame-to-frame change), c.) Blink rate and saccade rate within the interval (Alshanskaia et al., 2024; Grabinger et al., 2024).

### 3.3. Mouse tracking and interaction logging

Also, we continuously track the following mouse activity events, such as *mean movement velocity*, *path curvature*, *click frequency*, and *inter-click dwell times*. *Text selection events* are also tracked by their count and mean selection length, capturing copying behavior that occurs in case of comprehension difficulties.

### 3.4. Per-slide feature vector

The studying content (Programming language course for Python) is presented in form of slides with small snippets of code illustrating examples. For each slide, we track all physiological and behavioral events with their timestamps, which makes it easier to compare results for future evaluation. For each slide, we construct the following multimodal feature vector comprising:

A) **Eye-based features:** i.) Mean pupil diameter minus session baseline (pupil delta); ii.) Pupil velocity and variance; iii.) Number of fixations and fixation entropy (distribution of fixation durations); iv.) Blink rate and saccade rate (events per second).

B) **Mouse-based features:** i.) Mean and standard deviation of mouse movement velocity; ii.) Path curvature (ratio of traversed distance to direct distance); iii.) Click count and click frequency (clicks per second); iv.) Mean and variance of dwell time between clicks; v.) Number and mean length of text selections.



Figure 1. System Screenshot

### 3.5. Realtime cognitive load assessment

For each slide we group all feature vectors (eye and mouse metrics) into a single matrix, normalize each feature (using z-scores), and run lightweight k-means clustering with three clusters ( $k=3$  for *low*, *medium*, and *high* CL respectively). Clustering is performed separately for each user session, therefore the three groups identified each time reflect relative differences in CL for that learner and material, rather than relying on fixed thresholds.

For a reasonable interpretation of clusters (*labeling*) between *low*, *medium*, or *high* CL, we use **baseline-corrected pupil size** for each cluster to order the clusters from smallest to largest. This ordering builds on the well-validated indicator for higher cognitive effort. In the next step we map these three ranked clusters to numeric scores on a [0–10] scale, where low CL  $\approx 2$ , medium CL  $\approx 5$ , high CL  $\approx 8$  and track the computed results for each slide saving as CSV file. After each slide we also ask students on a scale 1-9 to assess subjectively the difficulty of each slide.

### 3.6. Word-level gaze analysis and local difficulty estimation

To even more fine-grained indicators of regional content difficulty, we record the spatial layout of each slide at the level of individual words. For each slide we compute bounding boxes for all rendered word tokens. During presentation, each gaze sample from the observer is tested whether it is within of any word bounding boxes for the currently available slide. When the gaze falls within a box, we increase that word's hit count and update its first/last hit time in a log file. After the slide is hidden (in our case it is either switching to a new slide or clicking "Help" button), we compute, for each word, a dwell time proxy as the difference between its first and last gaze timestamps. From these word-level measures, we derive an interpretable word-level cognitive load index on a 0–10 scale, combining normalized dwell time and hit count (longer dwell and more revisits imply higher difficulty), while filtering stop words and collapsing inflected forms.

### 3.7. LLM-based cognitive-aware learner adaptive hints

In our system we offer on-request (a help button), gaze- and load-aware learning support via a LLM (*gpt-oss:20b*). Using the following context for each slide (a) the current slide context, (b) a summary of overall CL with its interpretation and (c) a ranked list of difficult words derived from the word-level dwell index (0–10) described in the previous subsection, we send a structured prompt to the LLM, which generates a short slide summary, explanations of the most difficult terms, and 2–3 self-assessment questions for a self-reflection that are rendered in an popup over the slide. To prevent over-assistance and high extraneous CL, LLM help is automatically suggested only for those slides whose current CL is at least within *medium* level; for *low*-load slides, help remains available but is not promoted. In this case the user will be offered a summary when switching to a new slide (Santhosh et al., 2024).

#### 4. Conclusion and Future Work

Our paper presents a concept which has already been developed, and the early proof of concept was confirmed by several participants. For extended evaluation of the framework, it is planned to use the format of target and control group, where both groups will acquire a test assessing the baseline knowledge before viewing the learning materials, and after to compare the learning gains when using traditional and adaptive setup.

#### References

- Alshanskaia, E. I., Portnova, G. V., Liaukovich, K., & Martynova, O. V. (2024). Pupillometry and autonomic nervous system responses to cognitive load and false feedback: An unsupervised machine learning approach. *Frontiers in Neuroscience*, 18. <https://doi.org/10.3389/fnins.2024.1445697>
- Asteriadis, S., Tzouveli, P., Karpouzis, K., & Kollias, S. (2009). Estimation of behavioral user state based on eye gaze and head pose—Application in an e-learning environment. *Multimedia Tools and Applications*, 41(3), 469–493. <https://doi.org/10.1007/s11042-008-0240-1>
- Ayres, P., Lee, J. Y., Paas, F., & van Merriënboer, J. J. G. (2021). The Validity of Physiological Measures to Identify Differences in Intrinsic Cognitive Load. *Frontiers in Psychology*, 12, 702538. <https://doi.org/10.3389/fpsyg.2021.702538>
- de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, 38(2), 105–134. <https://doi.org/10.1007/s11251-009-9110-0>
- Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. (1988). In *Advances in Psychology* (Bd. 52, S. 139–183). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Duran, R., Zavgorodniaia, A., & Sorva, J. (2022). Cognitive Load Theory in Computing Education Research: A Review. *ACM Trans. Comput. Educ.*, 22(4), 40:1-40:27. <https://doi.org/10.1145/3483843>
- Grabinger, L., Hauser, F., Wolff, C., & Mottok, J. (2024). On Eye Tracking in Software Engineering. *SN Comput. Sci.*, 5(6). <https://doi.org/10.1007/s42979-024-03045-3>
- Herbig, N., Düwel, T., Helali, M., Eckhart, L., Schuck, P., Choudhury, S., & Krüger, A. (2020). Investigating Multi-Modal Measures for Cognitive Load Detection in E-Learning. *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 88–97. <https://doi.org/10.1145/3340631.3394861>
- Jabbour, J., Kleinbard, K., Miller, O., Haussman, R., & Reddi, V. J. (2025). *SocratiQ: A Generative AI-Powered Learning Companion for Personalized Education and Broader Accessibility* (arXiv:2502.00341). arXiv. <https://doi.org/10.48550/arXiv.2502.00341>
- Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence*, 2, 100017. <https://doi.org/10.1016/j.caeai.2021.100017>

Kaduk, T., Goeke, C., Finger, H., & König, P. (2024). Webcam eye tracking close to laboratory standards: Comparing a new webcam-based system and the EyeLink 1000. *Behavior Research Methods*, 56(5), 5002–5022. <https://doi.org/10.3758/s13428-023-02237-8>

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01997>

Leppink, J. (2017). Cognitive load theory: Practical implications and an important challenge. *Journal of Taibah University Medical Sciences*, 12(5), 385–391. <https://doi.org/10.1016/j.jtumed.2017.05.003>

Martin, F., Chen, Y., Moore, R., & Westine, C. (2020). Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018. *Educational Technology Research and Development*, 68, 1903–1929. <https://doi.org/10.1007/s11423-020-09793-2>

Paas, F. G. W. C., Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of Cognitive Load in Instructional Research. *Perceptual and Motor Skills*, 79(1), 419–430. <https://doi.org/10.2466/pms.1994.79.1.419>

PELICAN: Personalized Education via LLM-powered Cognitive Diagnosis and Adaptive Tutoring. (2025). <https://openreview.net/forum?id=NlvqjJ8R4J>

Santhosh, J., Dengel, A., & Ishimaru, S. (2024). Gaze-Driven Adaptive Learning System With ChatGPT-Generated Summaries. *IEEE Access*, 12, 173714–173733. <https://doi.org/10.1109/ACCESS.2024.3503059>

Santini, T., Fuhl, W., & Kasneci, E. (2018). PuRe: Robust pupil detection for real-time pervasive eye tracking. *Computer Vision and Image Understanding*, 170, 40–50. <https://doi.org/10.1016/j.cviu.2018.02.002>

Shemshack, A., & Spector, J. M. (2020). A systematic literature review of personalized learning terms. *Smart Learning Environments*, 7(1), 33. <https://doi.org/10.1186/s40561-020-00140-9>

Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G., & Vasileva, T. (2025). AI-driven adaptive learning for sustainable educational transformation. *Sustainable Development*, 33(2), 1921–1947. <https://doi.org/10.1002/sd.3221>

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>

The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload. (1988). In *Advances in Psychology* (Bd. 52, S. 185–218). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62387-0](https://doi.org/10.1016/S0166-4115(08)62387-0)

Wang, X., Huang, R. “Tammy”, Sommer, M., Pei, B., Shidfar, P., Rehman, M. S., Ritzhaupt, A. D., & Martin, F. (2024). The Efficacy of Artificial Intelligence-Enabled Adaptive Learning Systems From 2010 to 2022 on Learner Outcomes: A Meta-Analysis. *Journal of Educational Computing Research*, 62(6), 1348–1383. <https://doi.org/10.1177/07356331241240459>

Zagermann, J., Pfeil, U., & Reiterer, H. (2016). *Measuring Cognitive Load using Eye Tracking Technology in Visual Computing*. <https://kops.uni-konstanz.de/handle/123456789/38076>