# Studying Webcam-based Gaze Estimation and Mouse Coordination for Cognitive Inferences

Anas Doubabi
a.doubabi.ced@uca.ac.ma
Systems Engineering and Applications Laboratory (LISA), ENSA, Cadi Ayyad University
Marrakech, Morocco

Kenan Bektaş
kenan.bektas@unisg.ch
Institute of Computer Science, University of St. Gallen
St Gallen, Switzerland

Ahmed Aamouche
a.aamouche@uca.ac.ma
Systems Engineering and Applications Laboratory (LISA), ENSA, Cadi Ayyad University
Marrakech, Morocco

Hamada El Kabtane
h.elkabtane@uca.ac.ma
Research Laboratory in Intelligent and Sustainable Technologies, ENSA, Cadi Ayyad University
Marrakech, Morocco

Amine Abbad-Andaloussi
amine.abbad-andaloussi@unisg.ch
Institute of Computer Science, University of St. Gallen
St Gallen, Switzerland

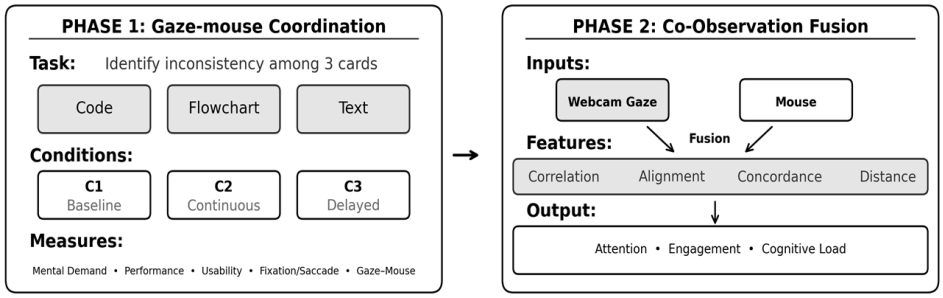


**Figure 1: Two phases of our Gaze–Mouse Co-Observation study for cognitive state inference.**

## ABSTRACT

Adaptive e-learning systems require scalable signals for inferring learners' cognitive states (e.g., attention, engagement, and cognitive load), yet webcam-based gaze estimation remains sensitive to calibration demands and real-world performance degradation. Mouse input is ubiquitous and calibration-free; however, cursor trajectories may only weakly reflect moment-to-moment visual attention. This work presents a preliminary research design that pairs a mouse-contingent blur interaction with a co-observation modeling view of cognitive state to make gaze–mouse data more useful under realistic constraints. First, we propose a mouse-contingent blur paradigm (i.e., *delayed blur after mouse inactivity*) and compare it with *no blur* and *mouse-contingent blur* to study how they affect gaze–mouse coordination and usability. Second, we frame webcam-based eye tracking and mouse input as cross-modal observations and motivate their fusion as a practical strategy to assess the changes in cognitive state of learners.


*ETRA '26, June 01–04, 2026, Marrakesh, Morocco*

ACM ISBN 979-8-4007-2519-7/2026/06.
https://doi.org/10.1145/3797246.3805849

## CCS CONCEPTS

• **Human-centered computing** → **Interaction paradigms**; *Usability testing*.

**ACM Reference Format:**
Anas Doubabi, Kenan Bektaş, Ahmed Aamouche, Hamada El Kabtane, and Amine Abbad-Andaloussi. 2026. Studying Webcam-based Gaze Estimation and Mouse Coordination for Cognitive Inferences. In *2026 Symposium on Eye Tracking Research and Applications (ETRA '26), June 01–04, 2026, Marrakesh, Morocco.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3797246.3805849

## 1 INTRODUCTION AND RELATED WORK

An adaptive e-learning system would benefit from a precise assessment of the learner's cognitive states, including attention, engagement, and cognitive load, to personalize content and improve learning outcomes [Zhou et al. 2017]. In practice, cognitive states are typically inferred through behavioral proxies of attention, yet continuous, scalable measurement remains challenging because many attention-sensitive signals are hard to detect in uncontrolled environments [Smilek et al. 2006]. Eye tracking can capture measurable distinctions in visual attention and cognitive load across varied learning tasks [Jin et al. 2025]. However, eye-tracking devices are costly, making them difficult to deploy at scale across all learning environments. Webcam-based gaze estimation can infer the point

of gaze (PoG) on a screen, however, it is susceptible to changes in various conditions such as illumination, head pose, eye wear, and camera quality [Cheng et al. 2024]. As a result, gaze signals may shift, degrade, or become intermittently unavailable in real-world deployment, limiting their usefulness for robust cognitive-state inference. Mouse input is not a direct proxy for gaze, as learners often keep the mouse stationary while reading or reflecting [Liebling and Dumais 2014]. However, it is ubiquitously available, does not require per-user calibration, and can be considered as a complementary signal to gaze. Therefore, it has been used in previous interaction designs and modeling approaches (e.g., Fisheye Views [Furnas 1986] or Focus+Context Interfaces [Cockburn et al. 2008]) that can assume persistent overt attention [Bektaş 2018].

Mouse-contingent blur has been proposed as a scalable alternative to eye tracking by revealing a cursor-centered region while blurring the remainder of the stimulus. Prior paradigms (e.g., RFV [Jansen et al. 2003], BubbleView [Kim et al. 2017], MouseView.js [Anwyl-Irvine et al. 2021], TurkEyes [Newman et al. 2020]) validate this approach primarily for static image viewing [Bektaş 2018; Jiang et al. 2015]. In e-learning, however, learners naturally alternate between reading, thinking, and intermittent interaction [Matzen et al. 2021]; under these usage patterns, mouse-contingent viewing ties content visibility directly to cursor position, forcing continuous cursor movement to keep content revealed. While this yields strong alignment between mouse and gaze movements, it raises usability concerns, as existing approaches usually blur all content outside the cursor area, which deviate from typical e-learning interfaces. To overcome this limitation. We introduce a *delayed-blur* condition that keeps the display unblurred during active mouse interaction and applies blur only after a period of mouse inactivity. This design aims for better usability, approaching natural e-learning settings while still encouraging gaze–mouse coordination. [Zhu et al. 2023] study activity monitoring in *e-learning* by combining gaze and mouse data for multimodal activity recognition. They collect a gaze–mouse dataset across routine digital activities and show that modeling cross-modal interactions between gaze and mouse (i.e., their coordination) improves recognition performance compared to using either modality alone. However, their approach relies on high-fidelity eye tracking under controlled conditions and does not address shifted or degraded gaze signals typical of webcam-based deployment.

Our work targets this gap by highlighting gaze–mouse coordination as a robust behavioral signal that can complement noisy or intermittently unavailable webcam-based gaze in naturalistic settings and encourages treating cross-modal coordination patterns as evidence for attention and engagement.

## 2 RESEARCH VISION

This work makes two contributions for robust cognitive-state inference under realistic webcam constraints. First, we introduce a mouse-contingent blur paradigm to encourage tighter gaze–mouse coupling with low interaction overhead. We focus on a *delayed-blur* condition that activates only after mouse inactivity. Second, we treat gaze and mouse as two imperfect but complementary signals of the same underlying cognitive state, rather than using mouse input solely to recover or approximate gaze. Under this view, mouse behavior can serve as a stable backbone when gaze data becomes sparse and unreliable, and combining both signals provides a more temporally dense and resilient inferential basis for the localizations of learners' cognitive state changes. Building on these two components, we outline a two-phase research agenda toward scalable cognitive state inference (e.g., attention, engagement, and cognitive load) in e-learning, under minimal or imprecise gaze calibration constraints. The agenda targets settings where webcam-based gaze estimation is noisy, intermittently unavailable, or hard to calibrate robustly. The goal shifts from precise gaze coordinates to robust cognitive-state inference under degraded sensing.

### 2.1 Phase 1: Gaze–Mouse Coordination

Phase 1 evaluates whether interaction design can increase gaze–mouse coupling without imposing motor overhead that would disrupt authentic learning. Following our design review, we compare three display modes (Figure 1): C1 (Uniform/Baseline: no blur), C2 (Mouse-contingent: hover-to-reveal), and C3 (Delayed blur: blur after mouse inactivity), using C1 as the reference for both usability and coordination.

*2.1.1 Hypothesis.* Delayed blur (C3) increases gaze–mouse coupling relative to baseline (C1), while preserving usability closer to baseline than continuous blur (C2).

*2.1.2 Study Design.*

*Independent variables:* Display blur mode, Stimulus inconsistency type, Task difficulty.

*Dependent variables:* Perceived mental demand, Task performance, Usability, Fixation and saccade-based eye-tracking measures, Gaze–mouse distance (for alignment).

*Experimental setup:* All blur parameters (e.g., aperture radius and blur strength) are defined in units of visual angle to ensure perceptual consistency across different display sizes and viewing distances. The inactivity threshold will be calibrated during piloting, informed by literature on reading dwell times and attention shifts, to distinguish high reading engagement from avoidance behavior. Webcam-based eye tracking is recorded in parallel with mouse input throughout each task to capture both modalities under identical interaction and content conditions.

*Materials:* Each task presents three complementary representations of the same programming concept as separate "cards": a Python code snippet, a flowchart describing the program logic, and a textual explanation of the code (or a code prompt). The cards are presented simultaneously to support cross-referencing during problem solving.

*Task design:* The task objective is to identify an inconsistency between the three cards (e.g., a mismatch between the code and the flowchart, or between the textual explanation and the implemented logic). This design draws on established paradigms in multi-representational learning and code comprehension, where cross-referencing tasks are recognized as cognitively demanding and ecologically valid. It also induces goal-directed visual search and

comprehension behavior that are representative of authentic e-learning activities, while providing a consistent objective across tasks and conditions[Alemdag and Cagiltay 2018].

*Difficulty control:* Stimulus extent (e.g., text length, number of flowchart nodes, and lines of code) is used as a proxy for difficulty. To validate this assumption, text difficulty will also be matched using readability scores, flowchart and code complexity will be cross-validated using established structural complexity metrics [Abbad-Andaloussi 2023], and tasks will be pilot-tested to collect perceived mental effort ratings and baseline performance (accuracy and completion time). Final task sets will be selected so these validation measures are comparable across conditions, reducing the risk that observed effects are driven by unintended difficulty differences.

*Design constraints:* Tasks are constructed to minimize learning effects across trials.

## 2.2 Phase 2: Cognitive Inference

Phase 2 operationalizes the co-observation view (Figure 1): rather than using mouse to calibrate gaze, we treat both as cross-modal observations of a latent cognitive state and fuse them for robust inference. Participants complete e-learning tasks under the optimal blur condition from Phase 1 while webcam gaze and mouse are recorded. We extract coordination features (e.g., movement correlation, temporal alignment and relative distance) that capture gaze–mouse coupling rather than absolute positions. These features are then evaluated as predictors of attention, engagement, and cognitive load under progressive gaze degradation, testing whether a mouse-as-backbone strategy preserves inference quality when webcam gaze is noisy or unavailable.

# 3 OUTLOOK

We present this design to clarify the underlying assumptions and to motivate discussion of the experimental conditions and task design for validating gaze–mouse coordination, as well as the multimodal inference framework for cognitive state estimation under sparse and unreliable gaze conditions. As the study is currently at the design stage, this contribution is intended to invite discussion of the underlying assumptions before empirical validation begins. The paradigm draws on a combination of approaches explored in prior work, which motivates its overall feasibility. While the current study focuses on a single task family with skilled adult learners, the paradigm can generalize to screen-based learning tasks that combine other content formats. We anticipate that these considerations may be of interest to researchers working on practical sensing for interactive systems where precise gaze calibration is unavailable or unreliable.

# REFERENCES


Amine Abbad-Andaloussi. 2023. On the relationship between source-code metrics and cognitive load: A Systematic Tertiary Review. *Journal of Systems and Software* 198 (04 2023), 111619. https://doi.org/10.1016/j.jss.2023.111619

Ecenaz Alemdag and Kursat Cagiltay. 2018. A systematic review of eye tracking research on multimedia learning. *Computers & Education* 125 (06 2018). https://doi.org/10.1016/j.compedu.2018.06.023

Alexander Anwyl-Irvine, Thomas Armstrong, and Edwin Dalmaijer. 2021. MouseView.js: Reliable and valid attention tracking in web-based experiments using a cursor-directed aperture. *Behavior Research Methods* 54 (09 2021). https://doi.org/10.3758/s13428-021-01703-5

Kenan Bektaş. 2018. *Gaze-contingent geovisualization for level of detail management.* Ph. D. Dissertation. University of Zurich. https://www.zora.uzh.ch/handle/20.500.14742/154491

Yihua Cheng, Haofei Wang, Yiwei Bao, and Feng Lu. 2024. Appearance-Based Gaze Estimation With Deep Learning: A Review and Benchmark . *IEEE Transactions on Pattern Analysis & Machine Intelligence* 46, 12 (Dec. 2024), 7509–7528. https://doi.org/10.1109/TPAMI.2024.3393571

Andy Cockburn, Amy Karlson, and Ben Bederson. 2008. A Review of Overview+Detail, Zooming, and Focus+Context Interfaces. *ACM Comput. Surv.* 41 (12 2008). https://doi.org/10.1145/1456650.1456652

George Furnas. 1986. Generalized Fisheye Views. *ACM Sigchi Bulletin* 17 (04 1986), 16–23. https://doi.org/10.1145/22339.22342

Anthony Jansen, Alan Blackwell, and Kim Marriott. 2003. A tool for tracking visual attention: The Restricted Focus Viewer. *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc* 35 (03 2003), 57–69.

Ming Jiang, Shengsheng Huang, Juanyong Duan, and Qi Zhao. 2015. SALICON: Saliency in Context. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1072–1080. https://doi.org/10.1109/CVPR.2015.7298710

Xiaofu Jin, Yunpeng Bai, Lina Xu, Shuai Ma, Danqing Shi, Luwen Yu, and Mingming Fan. 2025. Decoding Cognitive Load: Eye-Tracking Insights into Working Memory and Visual Attention. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications (ETRA '25)*. Association for Computing Machinery, New York, NY, USA, Article 83, 7 pages. https://doi.org/10.1145/3715669.3725864

Nam Kim, Zoya Bylinskii, Michelle Borkin, Krzysztof Gajos, Aude Oliva, Fredo Durand, and Hanspeter Pfister. 2017. BubbleView: An Interface for Crowdsourcing Image Importance Maps and Tracking Visual Attention. *ACM Transactions on Computer-Human Interaction* 24 (11 2017), 1–40. https://doi.org/10.1145/3131275

Daniel Liebling and Susan Dumais. 2014. Gaze and mouse coordination in everyday work. *UbiComp 2014 - Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (09 2014), 1141–1150. https://doi.org/10.1145/2638728.2641692

Laura Matzen, Mallory Stites, and Zoe Gastelum. 2021. Studying visual search without an eye tracker: an assessment of artificial foveation. *Cognitive Research: Principles and Implications* 6 (12 2021). https://doi.org/10.1186/s41235-021-00304-2

Anelise Newman, Barry McNamara, Camilo Fosco, Yun Zhang, Pat Sukhum, Matthew Tancik, Nam Kim, and Zoya Bylinskii. 2020. TurkEyes: A Web-Based Toolbox for Crowdsourcing Attention Data. 1–13. https://doi.org/10.1145/3313831.3376799

Daniel Smilek, Elina Birmingham, David Cameron, Walter Bischof, and Alan Kingstone. 2006. Cognitive Ethology and Exploring Attention in Real-World Scenes. *Brain Research* 1080, 1 (2006), 101–119. https://doi.org/10.1016/j.brainres.2005.12.090

Yun Zhou, Tao Xu, Yanping Cai, Xiaojun Wu, and Bei Dong. 2017. Monitoring Cognitive Workload in Online Videos Learning Through an EEG-Based Brain-Computer Interface. 64–73. https://doi.org/10.1007/978-3-319-58509-3_7

Rongrong Zhu, Liang Shi, Yunpeng Song, and Zhongmin Cai. 2023. Integrating Gaze and Mouse Via Joint Cross-Attention Fusion Net for Students' Activity Recognition in E-learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (2023), 1–35. https://doi.org/10.1145/3610876