What Shapes Participant Data Quality? A Scoping Review and Case Study of Crowdsourced Webcam Eye Tracking in AI Interviews

KA HEI CARRIE LAU, Chair of Human-Centered Technologies for Learning, Munich Center for Machine Learning (MCML), Technical University of Munich, Germany
ENKELEJDA KASNECI, Chair of Human-Centered Technologies for Learning, Technical University of Munich, Germany

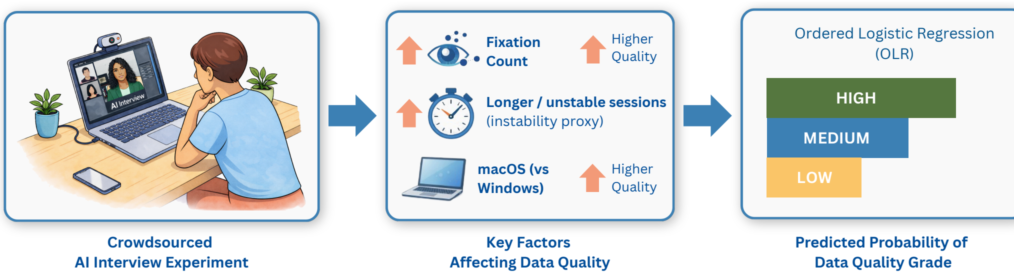


Fig. 1. Overview of our approach to evaluating participant data quality in crowdsourced webcam-based eye tracking during AI interviews. Behavioral and device-related factors predict data quality, and an ordered logistic regression (OLR) model is used to estimate the probability of different quality grades. Visuals were generated using GPT-5.3 and refined by the authors.

Webcam-based eye tracking is a cost-effective, scalable method for remote research that effectively reaches broader populations. However, uncontrolled environments and hardware diversity lead to inconsistent data quality in crowdsourcing. To assess current practices, we conducted a scoping review of crowdsourced eye-tracking from 2011–2025. The review confirms fragmented reporting and a lack of established quality benchmarks. To address this lack of predictive insight, we conducted a case study on AI fairness interviews ($N = 205$) using the RealEye platform. Applying Ordered Logistic Regression (OLR) to the platform's quality metric, we found that behavioral and technical factors significantly predict data quality. Specifically, within the RealEye platform, higher fixation counts, shorter sessions, and operating system choice yield significantly higher quality grades. Based on this review and platform-specific predictive insights, we provide actionable recommendations to enhance the reliability, transparency, and replicability of future crowdsourced webcam eye tracking in HCI and behavioral science.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; **Empirical studies in collaborative and social computing**.

Authors' Contact Information: Ka Hei Carrie Lau, carrie.lau@tum.de, Chair of Human-Centered Technologies for Learning, Munich Center for Machine Learning (MCML), Technical University of Munich, Munich, Germany; Enkelejda Kasneci, enkelejda.kasneci@tum.de, Chair of Human-Centered Technologies for Learning, Technical University of Munich, Munich, Germany.

ACM 2573-0142/2026/5-ARTETRA003
https://doi.org/10.1145/3806017

Additional Key Words and Phrases: eye tracking, webcam, crowdsourcing, data quality

ACM Reference Format:
Ka Hei Carrie Lau and Enkelejda Kasneci. 2026. What Shapes Participant Data Quality? A Scoping Review and Case Study of Crowdsourced Webcam Eye Tracking in AI Interviews. *Proc. ACM Hum.-Comput. Interact.* 10, 3, Article ETRA003 (May 2026), 19 pages. https://doi.org/10.1145/3806017

1 Introduction

Webcam-based eye tracking is a low-cost and scalable method that has democratized eye-tracking and behavioral research [James et al. 2025; Patterson et al. 2025; Prystauka et al. 2024; Van der Cruyssen et al. 2024; Vos et al. 2022; Yang and Krajbich 2021]. By leveraging participants' consumer webcams, it enables remote studies and access to diverse populations [Bertrand and Chapman 2023; Huang et al. 2018], extending the reach of research beyond the laboratory. Despite advances in webcam eye trackers, researchers still lack a systematic understanding of the reliability of crowdsourced webcam gaze data. Recent studies have reported inconsistent sampling precision and calibration stability across different hardware configurations [Heck et al. 2023; Patterson et al. 2025].

Unlike infrared (IR) eye trackers, webcam-based methods operate under uncontrolled conditions [Bánki et al. 2022; Semmelmann and Weigelt 2018]. Variations in lighting [Yang and Krajbich 2021], camera quality [Kim et al. 2017], hardware performance [Thilderkvist and Dobslaw 2024], and participant behavior [Kandel and Snedeker 2025] introduce noise that reduces spatial and temporal precision [Gagné and Franzen 2023; Juantorena et al. 2023; Thilderkvist and Dobslaw 2024]. Gaze estimates can lag by about 300 milliseconds (ms) relative to IR benchmarks [Slim and Hartsuiker 2023] and vary in sampling frequency or positional accuracy across devices [Ribeiro et al. 2023].

However, prior research has rarely quantified how participant behavior and device configuration affect data reliability. Individual studies have evaluated specific factors such as head movement [Sharafi et al. 2020; Thilderkvist and Dobslaw 2024] and screen distance [Juantorena et al. 2023], but few have statistically modeled these effects, particularly for commercial platforms like RealEye[1]. Without such a predictive framework, researchers cannot proactively identify or mitigate reliability issues in remote eye-tracking data, thereby limiting progress toward robust, replicable findings. Although our empirical analysis uses data provided by the RealEye platform, the factors we examine, such as fixation count, test duration, and operating system, are not specific to RealEye. Rather, they reflect general challenges of unsupervised, crowdsourced, webcam-based eye tracking, and our findings can help researchers understand the factors that influence data quality in such settings. To structure this research, we investigate the following research questions (RQs):

> **RQ1.** What methodological and validation practices have defined the development of webcam-based and crowdsourced eye-tracking research?
> **RQ2.** Which behavioral and technical factors predict data quality in crowdsourced webcam-based eye-tracking?

We address these questions through a scoping review of webcam-based and crowdsourced eye-tracking research from 2011 to 2025 and an empirical analysis of participant-level data quality. The review identifies three research areas: system development, validation, and application. Additionally, it highlights research gaps in quality reporting and predictive modeling. Our empirical analysis uses data from 205 participants in an artificial intelligence (AI) fairness interview study on the RealEye platform. Using ordered logistic regression (OLR), we investigated how behavioral and technical factors predicted the quality of gaze data. We chose this setting because AI interviews are a socially

[1]https://www.realeye.io/, last accessed 11 February 2026

interactive and attention-demanding task that mirrors realistic webcam-based interactions. This makes them a suitable context for evaluating data quality under unsupervised conditions. To our knowledge, this study is the first to combine a scoping review with an empirical, crowdsourced eye-tracking analysis in the emerging field of AI fairness.

Our findings show that webcam gaze data quality varies systematically rather than randomly. A higher total number of detected fixations per participant session and shorter test durations are associated with better data quality, and device-related factors also contribute significantly. These insights help refine guidelines for study design, participant screening, and data quality assessment, advancing methodological understanding of webcam-based eye tracking. To this end, we make two contributions:

Survey. We conduct a scoping review of crowdsourced webcam-based eye-tracking research across tasks, platforms, and validation methods. The review summarizes reported accuracy and data-loss measures, outlines current approaches, and identifies gaps in methodological transparency.

Empirical. We present empirical evidence on factors influencing data quality in crowdsourced webcam-based eye-tracking. Using OLR on a crowdsourced social perception dataset (N = 205), we evaluate the relationship between behavioral and technical variables and the likelihood of obtaining high-quality gaze data.

2 Related Work

Eye tracking has been applied across extended reality (XR), mobile, and webcam-based platforms, each of which poses distinct methodological challenges. Recent reviews highlight advances in gaze estimation, interaction design, and data analysis driven by low-cost sensors and deep learning [Adhanom et al. 2023; Bozkir et al. 2025; Katsini et al. 2020; Lei et al. 2023; Plopski et al. 2022]. However, they also note persistent issues with calibration reliability, data noise, and inconsistent reporting in unconstrained settings.

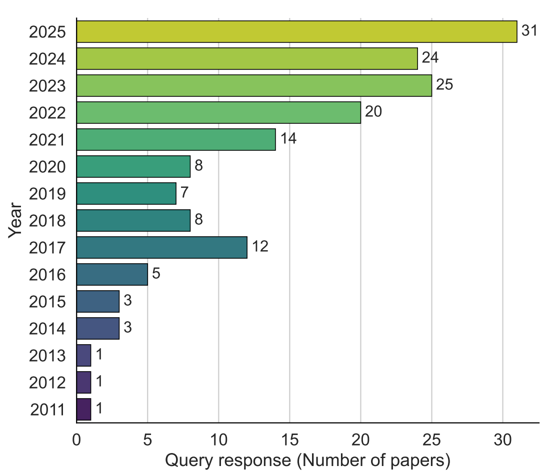
Methodological standards. Prior reviews have proposed guidelines for designing and reporting eye-tracking studies, emphasizing consistent calibration, transparent exclusion criteria, and standardized quality metrics [Blascheck et al. 2014; Carter and Luke 2020]. Building on this, Patterson et al. [2025] examined webcam-based and crowdsourced studies, identifying fragmented reporting and recommending structured documentation of sampling rates, calibration accuracy, and exclusion thresholds. Similar issues appear in XR and mobile research, where studies report trade-offs between scalability and precision and recurring concerns about privacy and robustness [Bozkir et al. 2025; Lei et al. 2023]. Altogether, these works underscore the need for standardized reporting frameworks to improve transparency and facilitate future cross-platform comparability.

System comparisons. Empirical comparisons between webcam-based and IR eye trackers reveal well-known issues, including spatial inaccuracy and calibration instability [Shehu et al. 2021]. Although deep learning has improved estimation robustness, webcam systems still face limitations in temporal precision and participant non-compliance [Patterson et al. 2025; Vos et al. 2022]. Most reviews remain descriptive, focusing on hardware or algorithmic performance rather than quantifying how behavioral or contextual factors influence data reliability.

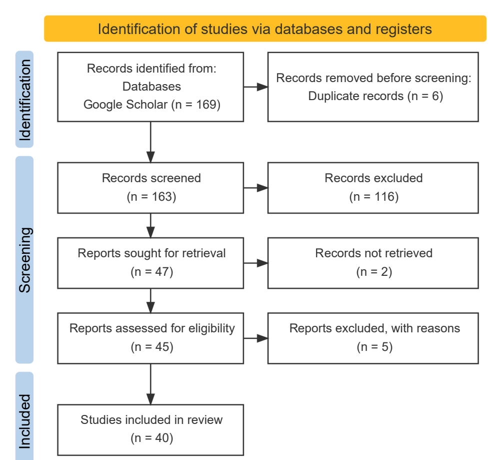
Research gap. Despite rapid methodological progress, validation and reporting practices have not kept pace. Existing reviews focus mainly on system accuracy and hardware performance, but rarely examine how behavioral and technical factors jointly shape webcam-based gaze data quality. As a result, reproducibility remains limited, and cross-platform benchmarks are still missing. To begin addressing this gap, we combine a scoping review with an empirical modeling analysis. The review summarizes current methods, validation strategies, and applications in webcam-based and crowdsourced eye tracking. The modeling analysis uses an existing RealEye dataset to explore how participant behavior and device factors relate to data quality, offering a first step toward more general, platform-independent models.

3 Scoping Review

To address **RQ1**, we searched Google Scholar for publications from 2011 to 2025. We used the query "crowdsourcing" AND ("eye tracking" OR "eye movement" OR "gaze") AND "webcam" AND ("data quality"). We did not restrict venues to avoid bias toward a single community, as webcam-based eye tracking spans human-computer interaction (HCI), psychology, computer vision, and marketing. We began with five survey papers as seed articles [Bozkir et al. 2025; Katsini et al. 2020; Patterson et al. 2025; Plopski et al. 2022; Shehu et al. 2021] and extended the set with backward and forward citation chasing. The database search identified 169 records. After removing duplicates (n=6), 163 unique records were screened as presented in Figure 2a, and 40 papers met the inclusion criteria. Figure 2b summarizes the identification, screening, and inclusion process. We included peer-reviewed, English-language studies on webcam-based or crowdsourced eye tracking that reported empirical findings or methodological evaluations. We excluded theses, editorials, position papers, tool-only notes, and non-English articles. Two researchers independently screened titles and abstracts and assessed full texts when necessary. Inter-rater reliability for screening decisions was substantial (Cohen's $\kappa = 0.78$). Disagreements were resolved through discussion, resulting in a final set of 40 included studies.



(a) Publications returned by the scoping review query (2011–2025).



(b) PRISMA flow diagram of study identification and screening.

Fig. 2. Overview of the scoping review process. (a) Number of publications returned by the query. After merging and removing duplicates, 163 unique records were screened and 40 studies were included. (b) PRISMA flow diagram of study identification, screening, and inclusion.

Our scoping review led to the identification of three main research directions that trace the development of webcam-based and crowdsourced eye-tracking research: (1) **Methodology**, which focuses on developing and improving the technology; (2) **Validity**, which establishes empirical trustworthiness through comparison with laboratory-based systems; and (3) **Application**, which applies webcam- and crowdsourced eye-tracking approaches to investigate behavioral, cognitive, and other research questions across different application domains.

3.1 Methodology

This line of research develops the core infrastructure for low-cost and webcam-based eye tracking. Across the literature, three major directions emerge: (1) **system and algorithm development**, (2) **dataset contribution**, and (3) **calibration and pipeline improvement**. These advancements reduce the costs and expertise required for traditional eye-tracking setups, enabling scalable studies to be conducted outside laboratory environments.

System and algorithm development. Methodological work primarily focuses on systems and algorithms that enable scalable webcam-based gaze tracking. There are two main approaches: **direct webcam eye tracking** and **attention proxy methods**. Early work that uses *direct webcam eye tracking* includes TurkerGaze [Xu et al. 2015], which demonstrated crowdsourced gaze collection via Amazon Mechanical Turk, while WebGazer.js [Papoutsaki et al. 2016] introduced an open-source, browser-based library for real-time gaze estimation using implicit calibration. WebGazer has since been integrated into experimental frameworks like jsPsych [James et al. 2025; Juantorena et al. 2023; Ribeiro et al. 2023; Vos et al. 2022; Yang and Krajbich 2021] and Gorilla [Bogdan et al. 2024; Prystauka et al. 2024]. SearchGazer [Papoutsaki et al. 2017] extended this approach for search tasks, improving drift correction through implicit interactions. Other webcam systems include UnitEye [Wagner et al. 2024] for 3D environments and Raspberry Pi–based deep-learning gaze predictors [Panja et al. 2025]. In contrast, *Attention Proxy Methods* such as BubbleView [Kim et al. 2017], FocalVid [Shaghaghi et al. 2025], and TurkEyes [Newman et al. 2020] approximate visual attention via mouse clicks or paths, providing scalable alternatives when webcam tracking is unavailable or privacy-sensitive.

Dataset contribution. A second key direction involves building datasets for training, validation, and benchmarking of remote eye-tracking and attention models. Such datasets support deep-learning development and enable performance evaluation across diverse hardware and participants. WebQAmGaze [Ribeiro et al. 2023] provides a multilingual webcam reading dataset collected with WebGazer and validated against EyeLink lab data. CrowdEyes [Othman et al. 2017] uses a low-cost head-mounted webcam and crowdsourcing via CrowdFlower to gather large-scale pupil-localization and fixation-tagging data, extending algorithm training beyond laboratory settings. BubbleView [Kim et al. 2017] and TurkEyes [Newman et al. 2020] similarly contribute crowdsourced attention maps and annotation frameworks, serving as both datasets and experimentation platforms. Altogether, these efforts establish the empirical foundation for testing and refining webcam-based gaze estimation methods.

Calibration and pipeline improvement. One crucial step toward robust webcam eye tracking is developing calibration methods that are both reliable and user-friendly. Recent research focuses on improving efficiency and personalization to mitigate the noise and variability inherent in real-world webcam data. For example, fast-PACE [Huang et al. 2018] builds on the Personalized Auto-Calibrating Eye-tracking (PACE) framework, which adapts gaze estimation automatically from natural user interactions such as clicks or typing, thereby reducing the need for explicit calibration. Saxena et al. [2022] evaluate streamlined calibration tasks, including brief pursuit routines and device-distance estimation, showing that shorter procedures can maintain accuracy while minimizing participant effort. Overall, these studies demonstrate that adaptive and lightweight calibration strategies are essential for making webcam-based eye tracking both accurate and practical beyond laboratory settings.

3.2 Validity

This line of research evaluates the empirical validity of webcam-based eye tracking by comparing its performance with laboratory-grade systems. Three main validation levels emerge: (1) **system-level**, assessing technical and measurement equivalence; (2) **task-level**, testing behavioral and cognitive replicability; and (3) **procedural-level**, defining best practices for reliable data collection under uncontrolled conditions.

System-level validation. System-level studies compare webcam-based eye trackers with laboratory-grade IR systems such as EyeLink and Tobii to assess technical performance [Asghari et al. 2022; Hammond and Wang 2023; Kaduk et al. 2024; Patterson et al. 2025; Slim and Hartsuiker 2023; Vos et al. 2022]. Three main limitations were identified: spatial inaccuracy, limited temporal precision, and systematic bias. Webcam eye trackers typically show spatial errors of about 3° to 4.5° [Asghari et al. 2022; Kaduk et al. 2024; Patterson et al. 2025; Vos et al. 2022], and sampling rates of 12–30 Hz restrict the detection of rapid eye movements compared with 100–1000 Hz in IR systems. They also exhibit centering bias, for example, gaze clustering near the screen center, and vertical compression, for instance, underestimation along the y-axis due to head motion, lighting variation, and geometric distortion [Kaduk et al. 2024; Slim and Hartsuiker 2023; Vos et al. 2022]. Although convolutional neural network (CNN)–based models can reduce errors to about 2.6° [Asghari et al. 2022], they cannot fully overcome the hardware and environmental constraints of consumer webcams. Nonetheless, strong correlations between webcam- and lab-based gaze trajectories, around $r = .8$ to .9 [Kaduk et al. 2024; Slim and Hartsuiker 2023; Vos et al. 2022], support webcam-based tracking for attentional and area-of-interest (AOIs) analyses. Overall, these findings define the technical boundaries within which webcam eye tracking produces valid behavioral data and motivate our analysis of factors predicting data quality in crowdsourced settings.

Task-level validation. This type of validation evaluates whether webcam-based eye tracking can reproduce well-known behavioral and cognitive effects. Across visual attention, language comprehension, and early cognitive development, webcam tracking replicates established gaze patterns, including predictive attention, novelty responses, and real-time language processing, though with smaller effect sizes (40–60% of laboratory results) and temporal delays of 200–700 ms [Bánki et al. 2022; Bogdan et al. 2024; Prystauka et al. 2024; Slim and Hartsuiker 2023; Swanson et al. 2024; Van der Cruyssen et al. 2024; Vos et al. 2022]. Van der Cruyssen et al. [2024] reproduced three classic gaze effects, and Vos et al. [2022] replicated verb-aspect processing with only a 50 ms delay. Emotion–attention [Bogdan et al. 2024] and infant looking-time [Bánki et al. 2022] studies confirm broader applicability despite lower spatial precision. Overall, webcam tracking yields valid behavioral measures when analyses target larger AOIs or sustained fixations.

Procedural-level validation. This level of research explores how procedural factors, including calibration routines, participant guidance, and recruitment quality, impact data reliability in uncontrolled online environments. Patterson et al. [2025] highlighted that transparent reporting of calibration thresholds, sampling-rate criteria, and participant instructions is essential for replicable webcam-based eye tracking. Likewise, Uittenhove et al. [2022] compared online and laboratory data collection and found that most data loss arises from participant non-compliance and sample quality rather than the testing environment itself. They reported that remote testing introduces only a small decrease in quality and recommended oversampling by about 20% while prioritizing participant screening procedures. These studies define procedural standards that improve the reliability and reproducibility of large-scale webcam-based eye-tracking research.

Despite this progress, it remains unclear which participant- and context-level factors most strongly predict data quality. Our study addresses this gap by modeling how behavioral and technical factors influence data reliability in crowdsourced webcam eye-tracking settings.

3.3 Applications

Research on webcam-based eye tracking demonstrates its use for studying attention, cognition, and behavior across three main domains: (1) **attention and interface studies**, which examine user engagement and visual saliency; (2) **cognitive and linguistic research**, which adapt classic experimental paradigms to online settings; and (3) **decision-making and behavioral economics**, which analyze how gaze dynamics influence attention and choice in complex decisions.

Attention and interface studies. A major application of webcam-based eye tracking is understanding how people attend to and interact with digital interfaces. Bertrand and Chapman [2023] examined gaze–cursor coordination during on-screen interaction, while Chen-Sankey et al. [2023] analyzed how young adults view e-cigarette marketing materials in realistic web environments. James et al. [2025] demonstrated that classic attention paradigms can be replicated online, supporting behavioral research despite spatial precision limits. In accessibility research, Edughele et al. [2022] reviewed gaze-based assistive systems that enable communication and interface control for individuals with motor impairments. Singh et al. [2023] introduced the multimodal *EngageNet* dataset to model user engagement in online learning, and Katsaounidou et al. [2025] developed the *iMedius* framework to monitor attention to online news and misinformation. Haveriku et al. [2025] further showed that eye-movement features enhance linguistic prediction and cross-lingual generalization. In sum, these studies demonstrate how webcam-based eye tracking supports diverse applications in interaction, accessibility, media, and language research.

Cognitive and linguistic research. Webcam-based eye tracking is also applied in cognitive and linguistic research. Juantorena et al. [2023] used a web-based prototype for the anti-saccade task, showing that inhibitory control and reaction-time effects can be measured reliably online. Thilderkvist and Dobslaw [2024] investigated how programmers read and comprehend source code, finding that gaze patterns and reading linearity differ from natural language and reveal distinct cognitive strategies. Yuksel Elgin and Elgin [2025] investigated how simulated visual field deficits affect information processing, showing that vision loss increases cognitive load and reduces comprehension through altered gaze behavior. These studies show that webcam-based methods enable remote investigation of executive control, comprehension, and information processing, extending cognitive and linguistic research beyond the laboratory.

Decision-making and behavioral economics. Webcam-based eye tracking reveals *how* people make decisions by capturing gaze dynamics as choices unfold. Yang and Krajbich [2021] found that longer and more frequent fixations predict choices and signal decision conflict, supporting the Attentional Drift Diffusion Model (aDDM), which posits that attended information receives greater weight. Bertrand et al. [2023] reported that harder choices elicit longer viewing times and more dwells. Wong [2023] showed that positive versus negative framing of supplier quality shifts attention and purchasing decisions, with gaze mediating this effect. Similarly, Sarvi et al. [2025] observed that visually distinct items attract earlier and longer fixations, linking saliency to consumer preference.

While prior application research has used webcam-based eye tracking across behavioral domains, few studies have examined participant-level data quality in these settings. Using data from an AI interview experiment, we model behavioral and technical factors that predict webcam gaze reliability in real-world conditions.

4 Case Study: Fairness in AI Interviews

To address **RQ2**, we analyzed a crowdsourced webcam eye-tracking dataset collected during AI-based job interviews [Lau et al. 2026]. The original study focused on participants' trust and fairness perceptions, whereas our goal here is to evaluate the reliability of RealEye's webcam-based tracking in this socially interactive and unsupervised setting. We model participant data quality by relating RealEye's quality grade to behavioral and technical factors recorded during the task.

4.1 Participants

The final sample comprised 205 valid datasets after excluding incomplete sessions and technical errors from 228 recruited individuals. Participants were adults fluent in English and located primarily in the United States, the United Kingdom, and Germany. All participants self-reported having normal or corrected-to-normal vision. The sample was demographically diverse (mean age $\approx$ 40 years) and representative of the typical heterogeneity of online crowdsourcing studies. No significant demographic differences were observed across participant quality grades. Detailed participant demographics are provided in Appendix Table 2.

4.2 Dataset and Procedure

We employed a 2×2 between-subjects design, manipulating the match or mismatch between participants' identities and the AI interviewer avatar by race and sex. We chose these avatar categories to reflect a majority–minority contrast motivated by evidence that hiring discrimination is associated with salient visual identity cues such as skin color, which signal perceived cultural distance [Zschirnt and Ruedin 2016]. Participants were recruited via Prolific [Prolific 2026b], between 2–17 July 2025. Eligibility required English fluency, a functioning webcam and microphone, and stable internet access. Each session lasted approximately 20 minutes, and participants were compensated £4.27 (= £12.80/hour) in accordance with Prolific's fair-pay policy [Prolific 2026a]. The overall study workflow, including recruitment, calibration, AI interview, and debriefing, is summarized in Figure 3. All procedures were approved by the Institutional Review Board (IRB) of the Technical University of Munich, and participants provided informed consent before participating.

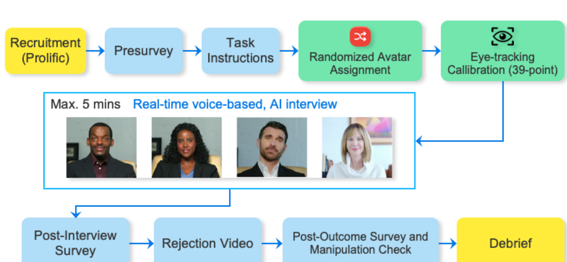


Fig. 3. Experimental procedure from recruitment to AI interview.

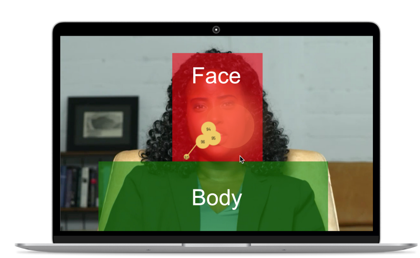


Fig. 4. Face and body AOIs defined on the AI interviewer.

4.3 Webcam Eye-Tracking Setup

Eye-tracking data were collected using the RealEye platform (version 18.49.0), a browser-based system that estimates gaze position from webcam video at a nominal sampling rate of 10–60 Hz, depending on device performance. The platform performs a 39-point calibration followed by a 3-point validation; participants who failed validation twice were automatically exited and compensated. According to RealEye's documentation, validation is passed only if the estimated gaze point falls within 150 pixels of each of the three validation targets. RealEye documentation also shows full-screen accuracy is approximately 100–125 pixels, with the highest accuracy in the central region. RealEye does not perform mid-task recalibration using a repeated calibration grid during the study session. Instead, the platform uses a "Virtual Chinrest" mechanism to maintain data quality: if a participant moves too far from the calibrated position, they are prompted to return to the correct position before tracking resumes. At present, RealEye does not provide an exportable metric for how many times the virtual chinrest was triggered during a session, which limits our ability to quantify tracking stability at the participant level.

We integrated RealEye's embedded SDK [RealEye 2026] to define two AOIs corresponding to the interviewer's face and body as shown in Figure 4. We defined only these two regions as AOIs because they are the main socially relevant targets in an interview context. We did not define a background AOI because background content varies across stimuli and is not central to our research questions. To mitigate the known limitations of webcam-based eye trackers in spatial precision and fixation accuracy as mentioned in [Patterson et al. 2025; Semmelmann and Weigelt 2018; Sharafi et al. 2020], the AOIs were defined large enough to capture all relevant fixations.

4.4 Measures and Analysis

The outcome variable was the **Participant Quality Grade** (1 = Very Low to 6 = Perfect), which RealEye computes from four internal signal metrics: sampling rate (Hz), fixation detection (i.e., a binary indicator of whether the system can compute fixations, requiring 20 Hz sampling rate), eye-tracking data length (also referred to as data integrity), which reflects the extent of gaps in the collected gaze data, and percentage of on-screen gaze time [RealEye 2025b]. Because RealEye does not provide per-participant gaze estimation error expressed in degrees of visual angle, we rely on the platform's quality grade as our primary outcome measure. Since these signals define the grade itself, we excluded them as predictors to avoid circularity. We instead used other measures provided by RealEye but not used in the grade calculation, along with information we collected separately.

Predictors were grouped into three categories: (1) *Behavioral factors*: fixation count and test duration (seconds), both exported by RealEye. Since fixation count is derived from event-detection procedures that identify fixations in continuous gaze streams [Kasneci et al. 2014], we interpret it here primarily as a behavioral proxy for participant engagement rather than a direct signal-quality metric. Fixation count comes from RealEye's event-detection pipeline but is not part of the grade formula; removing it worsened model fit by ΔAIC = 120, confirming its value as a predictor of engagement; (2) *Device factors*: browser width (px) and operating system, recorded separately; and (3) *Demographic factors*: participant age, when available in the RealEye metadata.

We report 95% Wald confidence intervals based on model standard errors. We tested the proportional-odds assumption using a partial proportional-odds (PPO) model that allowed non-parallel effects for *operating system*. The PPO did not improve fit (LR = 10.998, df = 8, p = .202; AIC_{PO} = 543.9 vs. AIC_{PPO} = 548.9), so we retained the proportional-odds model. Analyses were conducted in Python (version 3.9.6) and RStudio (version 2025.05.1+513).

4.5 Results: Quality Analysis

To examine how participant characteristics and session factors relate to data quality, we fitted an OLR model predicting participant quality grade. All predictors showed low multicollinearity (VIFs < 2), and model diagnostics indicated adequate fit (AIC = 519.9; McFadden's pseudo-R^2 = 0.212). Table 1 presents the full OLR results. The model identified four statistically significant predictors of participant quality grade: fixation count, test duration, browser width, and operating system. Participant age was not a significant predictor. We describe these results below, organized by behavioral factors, device-related factors, and operating system. To assess robustness, we ran an OLS model predicting mean sampling rate (Hz), reported in Appendix Table 3. The results confirmed the same direction of effects ($R^2 = 0.64$), indicating that the identified behavioral and device factors predict not only the platform's composite quality grade but also one of its constituent metrics, sampling rate. This convergence suggests that the identified predictors are not artifacts of RealEye's composite grading scheme.

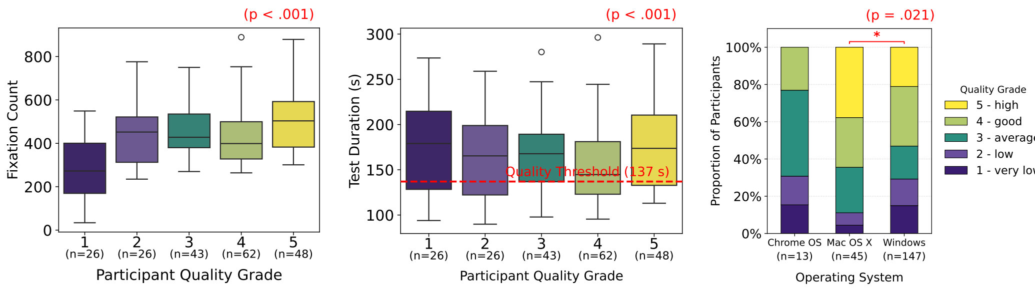
Table 1. Ordered logistic regression predicting participant quality grade. Positive coefficients indicate a higher likelihood of achieving a higher quality grade.

Variable	**Coef.**	**Std. Err.**	**z**	**P > |z|**	**[0.025]**	**[0.975]**
Fixation Count	0.0253***	0.003	8.527	0.000	0.019	0.031
Participant Age	-0.0175	0.012	-1.493	0.135	-0.040	0.005
Test Duration (s)	-0.0673***	0.009	-7.680	0.000	-0.084	-0.050
Test Browser Width (px)	0.0013**	0.000	3.054	0.002	0.000	0.002
Operating System (Mac OS X)	0.7358*	0.319	2.304	0.021	0.110	1.362
Cutpoints						
1/2	-1.5217	1.057	-1.440	0.150	-3.593	0.550
2/3	0.5509**	0.186	2.959	0.003	0.186	0.916
3/4	0.3343*	0.142	2.360	0.018	0.057	0.612
4/5	0.5427***	0.114	4.777	0.000	0.320	0.765
Model fit						
No. of observations						205
Log-Likelihood						-250.96
AIC						519.9
BIC						549.8
McFadden's pseudo R^2						0.212
Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$						

Behavioral factors. Fixation count was positively associated with participant data quality, as shown in Table 1 and Figure 5a. Each additional fixation increased the odds of belonging to a higher quality grade by about 2.5% (OR $\approx$ 1.025), holding other variables constant. In contrast, longer test durations were associated with lower quality grades, as shown in Table 1 and Figure 5b. Using the fitted model, we estimated a quality threshold, defined as the test duration at which a participant with average covariates has a 50% predicted chance of receiving a low-quality grade (Grade $\leq$ 3). The estimated threshold is 137 seconds, or about 46% of the 5-minute eye-tracking period. This value reflects the specifics of our setup and should be viewed as contextual guidance rather than a universal cut-off. The full session, including the interview, pre- and post-interview phases, and the post-outcome questionnaire, lasted about 20 minutes.

Device-related factors. Browser width was positively associated with participant quality grade, as shown in Table 1. Although the effect size is small, this association likely reflects aspects of display geometry and viewing conditions specific to the experimental setup (e.g., differences between desktop monitors and smaller laptop displays).

Operating system. The operating system was associated with participant quality grade, as shown in Table 1 and Figure 5c. Compared to the *Windows* baseline, *Mac OS X* users had higher odds of achieving a better quality grade.



(a) Fixation count (b) Test duration (c) Operating system

Fig. 5. Associations between key predictors and participant quality grade. (a) Fixation count by quality grade. (b) Test duration by quality grade; dashed line indicates the predicted 50% low-quality threshold (≈137 s). (c) Quality grade distribution by operating system, (*, p = .021), asterisks indicate the significant pairwise comparison. Inferential statistics are reported in Table 1.

5 Discussion

5.1 Summary of Findings

Our study addressed two research questions through a scoping review and an empirical quality analysis. **RQ1** explored the methodological and validation practices that have influenced webcam-based and crowdsourced eye-tracking research. The scoping review identified three main areas: system development, validation, and application. It showed that, although systems, algorithms, and datasets have advanced rapidly, validation and reporting practices have not kept pace. Few studies report standardized quality metrics or benchmarks, which limits reproducibility across platforms. Most validation work still targets system-level accuracy rather than participant or contextual factors, and procedural standards are inconsistently reported. These gaps highlight the need for predictive modeling approaches, such as this work, that link behavioral and technical variability to measurable data quality. In addressing **RQ2**, the regression analysis identified behavioral and device-related factors that significantly influenced webcam-based eye-tracking reliability, while demographic variables such as age showed no significant effect.

Behavioral factors. Our results show that *higher fixation count* predicted higher-quality grades, suggesting that attentive and consistent viewing behavior supports more reliable tracking. This finding is consistent with previous reading-based datasets, such as WebQAmGaze [Ribeiro et al. 2023], in which fixation density predicts comprehension accuracy. It is also consistent with our previous task-level validation results, which show that webcam-based eye tracking is most reliable for large AOIs and sustained fixations [Prystauka et al. 2024; Vos et al. 2022]. Both our results and those of previous studies suggest that fixation count captures both attentional engagement and physical stability, two conditions essential for accurate estimation in uncontrolled environments. In our study, all participants met RealEye's minimum sampling rate (20 Hz) for fixation detection, suggesting that variation in fixation count reflects differences in viewing behavior rather than detection failures.

Conversely, *longer test durations* predicted poorer quality. This pattern aligns with observations that extended testing often results from recalibration attempts or interruptions caused by head movement [Patterson et al. 2025; Vos et al. 2022]. While our dataset did not record the number of recalibration events, the RealEye platform includes a virtual chinrest feature that reminds participants to return to the calibrated head position when substantial head movement is detected [RealEye

2025a]. This mechanism may explain why longer test durations are associated with lower quality grades. In short, test duration acts as an indirect marker of participant fatigue and calibration instability rather than engagement per se.

Device-related factors. Data quality also varied systematically across participants' technical setups. *Wider browser windows* were associated with higher quality grades, as shown in Table 1. However, browser width in pixels does not account for physical screen size or pixel density, and may serve as a proxy for overall hardware and display quality (e.g., desktop monitors versus smaller laptops). Since the AOIs in our task were centrally positioned, this effect likely reflects the specific viewing conditions of our interview setup rather than a generalizable relationship between browser width and data quality. Operating-system differences supported prior system-level findings [Kaduk et al. 2024; Vos et al. 2022]. Participants using macOS achieved higher-quality grades, likely due to standardized camera drivers and consistent GPU timing that reduce sampling variability rather than inherent platform superiority [Kaduk et al. 2024; Vos et al. 2022]. Still, interactions between the operating system, browser, and hardware remain the main source of measurement differences in online eye tracking [Brandl et al. 2024; Prystauka et al. 2024; Uittenhove et al. 2022], and in some cases this pattern reverses: mobile users can yield more stable signals due to closer camera distance and steadier lighting [Chen-Sankey et al. 2023]. Overall, device effects depend on task context and environment, reflecting the interplay of the hardware–software ecosystem rather than any single platform.

Demographic factors. Participant demographics such as age showed no significant effect on data quality, consistent with prior webcam- and smartphone-based studies that found no demographic effects on data reliability [Bánki et al. 2022; Panja et al. 2025; Ribeiro et al. 2023]. Most data loss stems from non-compliance, inattention, or head movement rather than participant background [Bánki et al. 2022; Uittenhove et al. 2022].

5.2 Methodological Lessons

Conducting this study was our first experience combining crowdsourced recruitment with webcam-based eye tracking for a socially interactive AI interview. Inspired by Burch and Kurzhals [2024], we summarize our methodological lessons from two perspectives: that of **the researcher** and that of **the platform design**.

From the researcher's perspective. Our experience showed that conducting webcam-based eye-tracking studies with public participants presents both technical and procedural challenges. For many first-time users, calibration and sustained tracking required considerable effort. Several participants reported difficulties in the open-ended feedback, such as rapid head movements causing RealEye's virtual chinrest to reappear ("*The green dots part came up twice and took a long time to complete.*"). Such reports illustrate that while RealEye's virtual chinrest aims to improve data quality, it can also diminish the participant experience. Future studies should consider incorporating pre-study training or interactive calibration guidance, particularly when working with non-expert participants. For example, researchers may benefit from platforms that offer short animations illustrating appropriate lighting and posture, or simple real-time feedback to support participant self-correction during calibration.

From the platform design perspective. Some usability issues arose from interactions between the eye-tracking platform and our experimental interface. In some cases, RealEye's client-side interface elements interfered with task stimuli, e.g., one participant reported, ("*I had the pop-up for my eyes being in the circle a few times, and it blocked the exit button or covered the interviewer.*") Such conflicts highlight the need for better integration between eye-tracking platforms and experimental environments that rely on real-time rendering. Furthermore, greater transparency in how RealEye handles data loss would help researchers assess validity and reproducibility. Access to simple indicators, such as the number of times the virtual chinrest was triggered per participant, would provide valuable insight into tracking stability. We encourage providers to share additional quality metrics to support evidence-based methodological decisions.

5.3 Recommendations

Based on our findings and methodological reflections, we propose three recommendations for improving the design and success of crowdsourced webcam eye-tracking studies.

Recommendation 1: Provide clear setup and consent instructions. The amount of guidance participants receive directly influences data quality in webcam-based eye-tracking studies. As noted in previous work [Bánki et al. 2022; Patterson et al. 2025; Semmelmann and Weigelt 2018], researchers should provide clear, step-by-step instructions on device setup, lighting, and head stability, along with transparent explanations of data use and privacy protection. In our study, participants were informed that "*no video or audio recordings are stored, only gaze data are processed.*" and completed a short checklist to verify webcam functionality, internet stability, and calibration readiness. Clear preparation reduces tracking loss, ensures fair participation, and fosters trust in remote data collection. To support reproducibility, a concise reporting checklist for webcam-based eye-tracking studies is provided in the Appendix Table 4.

Recommendation 2: Evaluate platform architecture and integration early. Webcam eye-tracking platforms differ in how they handle data collection and calibration. Some platforms rely on screen recording or browser-based gaze estimation, while others embed experiments directly into their interfaces. These differences affect experimental control and data access. In our study, RealEye's embedded design limited control over calibration and interface rendering. Researchers should assess these constraints early to ensure the chosen platform aligns with their experimental goals.

Recommendation 3: Operationalize data quality for screening and exclusion. Researchers can operationalize data quality using behavioral indicators such as fixation count or test duration as post-hoc screening variables. For example, unusually long sessions or very low fixation counts may signal tracking instability or participant fatigue and can be flagged for sensitivity analyses or excluded based on pre-registered criteria. These predictors, identified in our analysis, provide practical guidance for defining quality thresholds in crowdsourced webcam-based eye-tracking studies.

5.4 Limitations and Future Work

Missing spatial accuracy metrics. RealEye does not provide per-participant gaze estimation error in degrees of visual angle, nor does it export raw validation offsets. Based on the platform's reported accuracy of approximately 106 pixels in the initial validation task [RealEye 2024] and assumptions about typical viewing distances (50–60 cm) and screen pixel densities (0.17–0.27 mm/px), we estimated spatial accuracy at 1.7° to 3.3°. Because neither viewing distance nor physical screen size were recorded, this estimate should be interpreted as an approximation rather than a participant-level measure. We encourage platform providers to report, for each participant, the mean gaze offset observed during validation, expressed in degrees of visual angle, to support independent quality assessment.

Data quality was not linked to task performance. We found no significant associations between quality grade and interview duration or transcript word count (Spearman, $N = 205$, all $p > .40$;

Kruskal–Wallis, all $p > .10$), though these proxies may not capture meaningful performance variation given the absence of well-defined accuracy outcomes. Prior work has shown that spatial inaccuracies can distort fixation-based measures such as dwell time [Brandl et al. 2024; Holmqvist et al. 2012], and that data quality moderates correlations between reading behavior and model predictions [Morger et al. 2022]. Future studies could correlate participants' attention check pass rates on the recruitment platform with their eye-tracking quality grades to assess whether recruitment-level screening predicts tracking reliability in crowdsourced settings.

Platform and task generalizability. The OLR results are specific to RealEye's algorithms and quality scoring system and may not transfer directly to other webcam-based platforms. Likewise, our AI interview task involved a limited set of participants' demographic and avatar identities. While these factors constrain generalization, the modeling approach provides a replicable template for identifying predictors of data quality across systems and study designs. Future work could test whether these predictors generalize across platforms and tasks, for example, by replicating the analysis with other commercial webcam eye-tracking services or task types (e.g., reading or visual search).

Sample characteristics. In our experiment, the distribution of quality grades was uneven, with few participants at the extremes. For the analysis, we merged Grades 5 and 6 due to the small highest-grade sample ($n = 3$). Model diagnostics confirmed adequate stability, but future research could ensure more balanced sampling across quality levels through stratified recruitment. Additionally, our study did not oversample to compensate for data loss due to calibration failures, a common issue in crowdsourced webcam eye-tracking. Following recent recommendations [Patterson et al. 2025], future studies could increase sample sizes by 20–40% beyond a priori power estimates.

6 Conclusion

This work combined a scoping review and a case study to evaluate factors influencing participant data quality in crowdsourced, webcam-based eye tracking. The review revealed fragmented reporting practices and limited consideration of behavioral and technical factors that affect data reliability. In our case study, analysis of data collected through the RealEye platform showed that fixation count, test duration, and operating system were consistent predictors of participant-level quality. These findings provide an empirical basis for assessing and improving remote gaze data. We encourage transparent reporting standards, reproducible analysis pipelines, and platform-independent quality models to enhance the reliability and comparability of webcam-based eye tracking in future research.

7 Societal Impact Statement

Webcam-based eye tracking offers scalable access to diverse participants but also raises concerns about privacy and data sovereignty. In this study, only gaze coordinates were recorded (no video), and all procedures followed IRB-approved consent and fair-compensation standards. We emphasize transparent data handling and advocate open reporting to promote privacy-conscious and ethical practices in future crowdsourced webcam-based eye-tracking research.

8 Open Science

To support reproducibility, we provide the complete analysis pipeline and scripts used for data processing at https://gitlab.lrz.de/hctl/crowdsourced-webcam-eyetracking-analysis.

References

Isayas Berhe Adhanom, Paul MacNeilage, and Eelke Folmer. 2023. Eye Tracking in Virtual Reality: a Broad Review of Applications and Challenges. *Virtual Reality* 27, 2 (01 Jun 2023), 1481–1505. doi:10.1007/s10055-022-00738-z

Parviz Asghari, Maike Schindler, and Achim J Lilienthal. 2022. Can eye tracking with pervasive webcams replace dedicated eye trackers? An experimental comparison of eye-tracking performance. In *International Conference on Human-Computer Interaction*. Springer, Cham, Switzerland, 3–10.

Anna Bánki, Martina De Eccher, Lilith Falschlehner, Stefanie Hoehl, and Gabriela Markova. 2022. Comparing online webcam-and laboratory-based eye-tracking for the assessment of infants' audio-visual synchrony perception. *Frontiers in psychology* 12 (2022), 733933.

Jennifer K Bertrand and Craig S Chapman. 2023. Dynamics of eye-hand coordination are flexibly preserved in eye-cursor coordination during an online, digital, object interaction task. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 517, 13 pages. doi:10.1145/3544548.3580866

Jennifer K. Bertrand, Alexandra A. Ouellette Zuk, and Craig S. Chapman. 2023. Continuous Measures of Decision-Difficulty Captured Remotely: II. Webcam eye-tracking reveals early decision processing. arXiv:https://www.biorxiv.org/content/early/2023/06/07/2023.06.06.543799.full.pdf doi:10.1101/2023.06.06.543799

Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. 2014. State-of-the-art of visualization for eye tracking data.. In *Eurovis (stars)*. Eurographics Association, Goslar, Germany, 29.

Paul C Bogdan, Sanda Dolcos, Simona Buetti, Alejandro Lleras, and Florin Dolcos. 2024. Investigating the suitability of online eye tracking for psychological research: Evidence from comparisons with in-person data using emotion–attention interaction tasks. *Behavior Research Methods* 56, 3 (2024), 2213–2226.

Efe Bozkir, Süleyman Özdel, Mengdi Wang, Brendan David-John, Hong Gao, Kevin Butler, Eakta Jain, and Enkelejda Kasneci. 2025. Eye-Tracked Virtual Reality: A Comprehensive Survey on Methods and Privacy Challenges. *Proc. IEEE* 113, 10 (2025), 1155–1191. doi:10.1109/JPROC.2026.3653661

Stephanie Brandl, Oliver Eberle, Tiago Ribeiro, Anders Søgaard, and Nora Hollenstein. 2024. Evaluating Webcam-based Gaze Data as an Alternative for Human Rationale Annotations.

Michael Burch and Kuno Kurzhals. 2024. Teaching Eye Tracking: Challenges and Perspectives. *Proc. ACM Hum.-Comput. Interact.* 8, ETRA, Article 237 (May 2024), 17 pages. doi:10.1145/3655611

Benjamin T. Carter and Steven G. Luke. 2020. Best practices in eye tracking research. *International Journal of Psychophysiology* 155 (2020), 49–62. doi:10.1016/j.ijpsycho.2020.05.010

Julia Chen-Sankey, Maryam Elhabashy, Stefanie Gratale, Jason Geller, Melissa Mercincavage, Andrew A Strasser, Cristine D Delnevo, Michelle Jeong, Olivia A Wackowski, et al. 2023. Examining visual attention to tobacco marketing materials among young adult smokers: Protocol for a remote webcam-based eye-tracking experiment. *JMIR Research Protocols* 12, 1 (2023), e43512.

Hilary O Edughele, Yinghui Zhang, Firdaus Muhammad-Sukki, Quoc-Tuan Vien, Haley Morris-Cafiero, and Michael Opoku Agyeman. 2022. Eye-tracking assistive technologies for individuals with amyotrophic lateral sclerosis. *IEEE Access* 10 (2022), 41952–41972.

Nathan Gagné and Léon Franzen. 2023. How to run behavioural experiments online: Best practice suggestions for cognitive psychology and neuroscience. *Swiss Psychology Open: the official journal of the Swiss Psychological Society* 3, 1 (2023).

Robert W Hammond and Yuqi Wang. 2023. Crowdsourced Online Biometric Studies: Is the juice worth the squeeze? *Muma Business Review* 7 (2023), 141–148.

Alba Haveriku, Hakik Paci, Nelda Kote, Paola Shasivari, and Elinda Kajo Meçe. 2025. A systematic review of eye-tracking data in NLP: exploring low-cost and cross-lingual possibilities. *International Journal of Grid and Utility Computing* 16, 1 (2025), 29–40.

Melanie Heck, Christian Becker, and Viola Deutscher. 2023. Webcam Eye Tracking for Desktop and Mobile Devices: A Systematic Review. In *Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS '23)*. University of Hawaii at Manoa, Honolulu, HI, USA, 1–10. doi:10.24251/HICSS.2023.825

Kenneth Holmqvist, Marcus Nyström, and Fiona Mulvey. 2012. Eye tracker data quality: what it is and how to measure it. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 45–52. doi:10.1145/2168556.2168563

Michael Xuelin Huang, Jiajia Li, Grace Ngai, and Hong Va Leong. 2018. Quick bootstrapping of a personalized gaze model from real-use interactions. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9, 4 (2018), 1–25.

Ariel N James, Rachel Ryskin, Joshua K Hartshorne, Haylee Backs, Nandeeta Bala, Laila Barcenas-Meade, Samata Bhattarai, Tessa Charles, Gerasimos Copoulos, Claire Coss, et al. 2025. What Paradigms Can Webcam Eye-Tracking Be Used For? Attempted Replications of Five Cognitive Science Experiments. *Collabra: Psychology* 11, 1 (2025), 140755.

Gustavo E Juantorena, Francisco Figari, Agustín Petroni, and Juan E Kamienkowski. 2023. Web-based eye-tracking for remote cognitive assessments: The anti-saccade task as a case study. 2023–07 pages. doi:10.1101/2023.07.11.548447

Tobiasz Kaduk, Caspar Goeke, Holger Finger, and Peter König. 2024. Webcam eye tracking close to laboratory standards: Comparing a new webcam-based system and the EyeLink 1000. *Behavior research methods* 56, 5 (2024), 5002–5022.

Margaret Kandel and Jesse Snedeker. 2025. Assessing two methods of webcam-based eye-tracking for child language research. *Journal of Child Language* 52, 3 (2025), 675–708.

Enkelejda Kasneci, Gjergji Kasneci, Thomas C. Kübler, and Wolfgang Rosenstiel. 2014. The applicability of probabilistic methods to the online recognition of fixations and saccades in dynamic scenes. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Safety Harbor, Florida) *(ETRA '14)*. Association for Computing Machinery, New York, NY, USA, 323–326. doi:10.1145/2578153.2578213

Anastasia Katsaounidou, Paris Xylogiannis, Thomai Baltzi, Theodora Saridou, Symeon Papadopoulos, and Charalampos Dimoulas. 2025. An AI-Driven News Impact Monitoring Framework Through Attention Tracking. *Societies* 15, 8 (2025).

Christina Katsini, Yasmeen Abdrabou, George E. Raptis, Mohamed Khamis, and Florian Alt. 2020. The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–21. doi:10.1145/3313831.3376840

Nam Wook Kim, Zoya Bylinskii, Michelle A Borkin, Krzysztof Z 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 (TOCHI)* 24, 5 (2017), 1–40.

Ka Hei Carrie Lau, Philipp Stark, Efe Bozkir, and Enkelejda Kasneci. 2026. Skin-Deep Bias: How Avatar Appearances Shape Perceptions of AI Hiring. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, Article 125, 20 pages. doi:10.1145/3772318.3790379

Yaxiong Lei, Shijing He, Mohamed Khamis, and Juan Ye. 2023. An end-to-end review of gaze estimation and its interactive applications on handheld mobile devices. *Comput. Surveys* 56, 2 (2023), 1–38.

Felix Morger, Stephanie Brandl, Lisa Beinborn, and Nora Hollenstein. 2022. A Cross-lingual Comparison of Human and Model Relative Word Importance. In *Proceedings of the 2022 CLASP Conference on (Dis)embodiment*, Simon Dobnik, Julian Grove, and Asad Sayeed (Eds.). Association for Computational Linguistics, Gothenburg, Sweden, 11–23. https://aclanthology.org/2022.clasp-1.2/

Anelise Newman, Barry McNamara, Camilo Fosco, Yun Bin Zhang, Pat Sukhum, Matthew Tancik, Nam Wook Kim, and Zoya Bylinskii. 2020. TurkEyes: A Web-Based Toolbox for Crowdsourcing Attention Data. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376799

Mohammad Othman, Telmo Amaral, Róisín McNaney, Jan D. Smeddinck, John Vines, and Patrick Olivier. 2017. CrowdEyes: crowdsourcing for robust real-world mobile eye tracking. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 18, 13 pages. doi:10.1145/3098279.3098559

Soumya Panja, Sapta Rathi Roy, Shatoparna Bhattacharya, Anshuman Kumar, and Debayan Bhattacharya. 2025. Prediction of Gaze Point Using Deep Learning and Raspberry Pi. In *Recent Advances in Artificial Intelligence and Smart Applications*, Jyotsna K. Mandal, Mike Hinchey, and Satyajit Chakrabarti (Eds.). Springer Nature Singapore, Singapore, 99–108.

Alexandra Papoutsaki, James Laskey, and Jeff Huang. 2017. SearchGazer: Webcam Eye Tracking for Remote Studies of Web Search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (Oslo, Norway) *(CHIIR '17)*. Association for Computing Machinery, New York, NY, USA, 17–26. doi:10.1145/3020165.3020170

Alexandra Papoutsaki, Patsorn Sangkloy, James Laskey, Nediyana Daskalova, Jeff Huang, and James Hays. 2016. Webgazer: scalable webcam eye tracking using user interactions. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16)*. AAAI Press, New York, New York, USA, 3839–3845.

Allie Spencer Patterson, Christopher Nicklin, and Joseph P Vitta. 2025. Methodological recommendations for webcam-based eye tracking: A scoping review. *Research Methods in Applied Linguistics* 4, 3 (2025), 100244.

Alexander Plopski, Teresa Hirzle, Nahal Norouzi, Long Qian, Gerd Bruder, and Tobias Langlotz. 2022. The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-worn Extended Reality. *ACM Comput. Surv.* 55, 3, Article 53 (March 2022), 39 pages. doi:10.1145/3491207

Prolific. 2026a. How Much Should You Pay Research Participants? https://www.prolific.com/resources/how-much-should-you-pay-research-participants. Accessed: 11 February 2026.

Prolific. 2026b. Prolific Crowdsourcing Platform. https://www.prolific.com/. Accessed: 11 February 2026.

Yanina Prystauka, Gerry TM Altmann, and Jason Rothman. 2024. Online eye tracking and real-time sentence processing: On opportunities and efficacy for capturing psycholinguistic effects of different magnitudes and diversity. *Behavior Research Methods* 56, 4 (2024), 3504–3522.

RealEye. 2024. *RealEye White Paper.* Technical Report. RealEye. Available upon request from https://www.realeye.io/lp/whitepaper.

RealEye. 2025a. How RealEye Works. https://support.realeye.io/how-realeye-works. Accessed: 3 November 2025.

RealEye. 2025b. Participant Quality Stats Explained. https://support.realeye.io/participant-quality-stats-explained. Accessed: November 3, 2025.

RealEye. 2026. Embedded Website SDK. https://app.realeye.io/docs/embedded-website-sdk. Accessed: 11 February 2026.

Tiago Ribeiro, Stephanie Brandl, Anders Søgaard, and Nora Hollenstein. 2023. WebQAmGaze: A Multilingual Webcam Eye-Tracking-While-Reading Dataset. arXiv:2303.17876 [cs.CL]

Fatemeh Sarvi, Mohammad Aliannejadi, Sebastian Schelter, and Maarten de Rijke. 2025. Understanding Visual Saliency of Outlier Items in Product Search.

Shreshth Saxena, Elke Lange, and Lauren Fink. 2022. Towards efficient calibration for webcam eye-tracking in online experiments. In *2022 Symposium on Eye Tracking Research and Applications* (Seattle, WA, USA) *(ETRA '22)*. Association for Computing Machinery, New York, NY, USA, Article 27, 7 pages. doi:10.1145/3517031.3529645

Kilian Semmelmann and Sarah Weigelt. 2018. Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods* 50, 2 (2018), 451–465.

Sahand Shaghaghi, Karissa B. Payne, Bryan Tripp, Kerstin Dautenhahn, and Chrystopher L. Nehaniv. 2025. FocalVid: A Platform for Tracking Visual Attention to Video via Crowdsourcing Validated Against Human Gaze Data. *IEEE Access* 13 (2025), 159566–159581. doi:10.1109/ACCESS.2025.3608621

Zohreh Sharafi, Bonita Sharif, Yann-Gaël Guéhéneuc, Andrew Begel, Roman Bednarik, and Martha Crosby. 2020. A practical guide on conducting eye tracking studies in software engineering. *Empirical Software Engineering* 25, 5 (2020), 3128–3174.

Ibrahim Shehi Shehu, Yafei Wang, Athuman Mohamed Athuman, and Xianping Fu. 2021. Remote eye gaze tracking research: A comparative evaluation on past and recent progress. *Electronics* 10, 24 (2021), 3165.

Monisha Singh, Ximi Hoque, Donghuo Zeng, Yanan Wang, Kazushi Ikeda, and Abhinav Dhall. 2023. Do I Have Your Attention: A Large Scale Engagement Prediction Dataset and Baselines. In *Proceedings of the 25th International Conference on Multimodal Interaction* (Paris, France) *(ICMI '23)*. Association for Computing Machinery, New York, NY, USA, 174–182. doi:10.1145/3577190.3614164

Mieke Sarah Slim and Robert J Hartsuiker. 2023. Moving visual world experiments online? A web-based replication of Dijkgraaf, Hartsuiker, and Duyck (2017) using PCIbex and WebGazer. js. *Behavior Research Methods* 55, 7 (2023), 3786–3804.

Elizabeth Swanson, Michael C Frank, and Judith Degen. 2024. Syntactic adaptation and word learning in children and adults. *Language Development Research* 4, 1 (2024), 1–41.

Enkelejda Tafaj, Thomas C. Kübler, Gjergji Kasneci, Wolfgang Rosenstiel, and Martin Bogdan. 2013. Online Classification of Eye Tracking Data for Automated Analysis of Traffic Hazard Perception. In *Artificial Neural Networks and Machine Learning – ICANN 2013*, Valeri Mladenov, Petia Koprinkova-Hristova, Günther Palm, Alessandro E. P. Villa, Bruno Appollini, and Nikola Kasabov (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 442–450.

Eva Thilderkvist and Felix Dobslaw. 2024. On current limitations of online eye-tracking to study the visual processing of source code. *Information and Software Technology* 174 (2024), 107502. doi:10.1016/j.infsof.2024.107502

Kim L Uittenhove, Stephanie Jeanneret, and Evie Vergauwe. 2022. From Lab-Testing to Web-Testing in Cognitive Research: Who You Test is More Important Than How You Test. doi:10.31234/osf.io/uy4kb

Ine Van der Cruyssen, Gershon Ben-Shakhar, Yoni Pertzov, Nitzan Guy, Quinn Cabooter, Lukas J Gunschera, and Bruno Verschuere. 2024. The validation of online webcam-based eye-tracking: The replication of the cascade effect, the novelty preference, and the visual world paradigm. *Behavior Research Methods* 56, 5 (2024), 4836–4849.

Myrte Vos, Serge Minor, and Gillian Ramchand. 2022. Comparing infrared and webcam eye tracking in the Visual World Paradigm. doi:10.31234/osf.io/36skd

Tobias Wagner, Mark Colley, Daniel Breckel, Michael Kösel, and Enrico Rukzio. 2024. UnitEye: Introducing a User-Friendly Plugin to Democratize Eye Tracking Technology in Unity Environments. In *Proceedings of Mensch Und Computer 2024* (Karlsruhe, Germany) *(MuC '24)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3670653.3670655

Ricky S Wong. 2023. An experimental investigation of attribute framing effects on risky sourcing behaviour: the mediating role of attention allocated to suppliers' quality information. *International Journal of Operations & Production Management* 43, 13 (2023), 205–225.

Pingmei Xu, Krista A Ehinger, Yinda Zhang, Adam Finkelstein, Sanjeev R Kulkarni, and Jianxiong Xiao. 2015. Turkergaze: Crowdsourcing saliency with webcam based eye tracking.

Xiaozhi Yang and Ian Krajbich. 2021. Webcam-based online eye-tracking for behavioral research. *Judgment and Decision making* 16, 6 (2021), 1485–1505.

Cansu Yuksel Elgin and Ceyhun Elgin. 2025. Visual Attention to Economic Information in Simulated Ophthalmic Deficits: A Remote Eye-Tracking Study. *Journal of Eye Movement Research* 18, 5 (2025), 50.

Eva Zschirnt and Didier Ruedin. 2016. Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. *Journal of Ethnic and Migration Studies* 42, 7 (2016), 1115–1134. arXiv:https://doi.org/10.1080/1369183X.2015.1133279 doi:10.1080/1369183X.2015.1133279

A Participant Demographics

Table 2. Participant demographics and individual differences by quality grade (N = 205). Values represent means ± standard deviations, or counts as indicated.

Quality Grade (n)	**Gender** (M/W)	**Age** (years)	**Ethnicity** (W/B)	**Vision** (N/G/C)	**Employment** (%)	**Interview** (Exp. %)	**Nervousness** (1–10)
1 (n=26)	10/16	42.2 ± 11.1	18/8	17/9/0	85	58	4.54 ± 2.76
2 (n=26)	17/9	42.2 ± 11.5	12/14	16/9/1	92	81	4.12 ± 2.82
3 (n=43)	19/24	38.3 ± 10.5	14/29	29/11/3	91	79	4.00 ± 2.47
4 (n=62)	28/34	40.9 ± 11.9	36/26	45/14/3	90	73	4.02 ± 2.55
5 (n=48)	28/20	37.1 ± 11.9	27/21	29/17/2	81	71	4.21 ± 2.77

Notes. Gender (M/W) = Men/Women; Ethnicity (W/B) = White/Black; Vision (N/G/C) = Normal/Glasses/Contact lenses; Employment values indicate the percentage of participants employed full- or part-time; Interview experience values indicate the percentage reporting moderate or extensive prior interview experience.

B Robustness Check

Table 3. OLS regression predicting mean sampling rate (Hz). The model uses the same predictors as the main OLR (Table 1) to assess robustness. Positive coefficients indicate higher sampling rates.

Variable	**Coef.**	**Std. Err.**	**t**	**P > |t|**	**[0.025]**	**[0.975]**
Fixation Count	0.0513^{***}	0.003	18.506	0.000	0.046	0.057
Participant Age	0.0100	0.023	0.427	0.670	-0.036	0.056
Test Duration (s)	-0.1354^{***}	0.009	-15.032	0.000	-0.153	-0.118
Test Browser Width (px)	0.0031^{***}	0.001	3.556	0.000	0.001	0.005
Operating System (Mac OS X)	0.8352	0.655	1.274	0.204	-0.457	2.128
Model fit						
No. of observations						205
Log-Likelihood						-561.91
AIC						1136
BIC						1156
R^2						0.642
Adj. R^2						0.633
Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$						

C Reporting Checklist

The following checklist summarizes essential reporting items for planning, conducting, and documenting crowdsourced webcam-based eye-tracking experiments.

Table 4. Checklist for conducting crowdsourced webcam-based eye-tracking studies.

Category	Checklist Item
Participant Setup	• Confirm webcam and browser compatibility before participation. • Provide clear setup instructions (lighting, posture, viewing distance). • Inform participants about data use and privacy. • Verify calibration readiness and camera permissions. • Screen for visual or hardware limitations that affect tracking.
Platform Configuration	• Specify the platform provider, version, and gaze estimation mode. • Describe calibration and validation procedures with pass/fail criteria. • Report nominal and measured sampling rates. • Indicate how recalibration or data loss is handled automatically.
Data Quality Metrics	• Define the quality indicator used (platform-provided or user-defined). • Report inclusion and exclusion thresholds and handling of missing data. • Report key signal metrics that contribute to data quality (e.g., sampling rate, valid-sample, gaze-on-screen).
Behavioral Measures	• Describe event-detection algorithms and parameters [Tafaj et al. 2013]. • Report session duration, gaze recording length, and engagement proxies. • Provide AOI definitions and any normalization method used.
Device and Context	• Record operating system and browser (name and version). • Record viewport or browser window dimensions (px). • Note display mode (fullscreen vs. windowed) and any zoom/scaling.
Demographics and Ethics	• Report demographic variables collected and screening criteria. • Describe informed consent, compensation, and data-retention policy. • State IRB or ethics approval reference.
Reproducibility	• Share code, data dictionary, and analysis scripts if permitted. • Note preprocessing or exclusion scripts used for quality control.