Bringing Sexy (Webcam Eye-tracking) Back into the lab: Stage 1 Registered Report

Jason Geller[1*], João Veríssimo[2] & Julia Droulin[3]

[1]Boston College, Department of Psychology and Neuroscience
[2]University of Lisbon, School of Arts and Humanities
[3]Univetsity of North Carolina - Chapel Hill, Speech and Hearing Divison

Webcam-based eye-tracking offers a scalable and accessible alternative to traditional lab-based systems. While recent studies demonstrate that webcam eye-tracking can replicate canonical effects across domains such as language, memory, and decision-making, questions remain about its precision and reliability. In particular, spatial accuracy, temporal resolution, and attrition rates are often poorer than those observed with research-grade systems, raising the possibility that environmental and hardware factors introduce substantial noise. The present registered report directly tests this hypothesis by bringing webcam eye-tracking back into a controlled laboratory setting. In Experiment 1, we examine the effect of webcam quality (high vs. low) in a single word Visual World Paradigm (VWP) task, testing whether higher-quality webcams yield stronger competition effects, earlier effect onsets, and reduced attrition. In Experiment 2, we assess the impact of head stabilization (chinrest vs. no chinrest) under identical environmental conditions. Together, these studies isolate the causal influence of hardware and movement on webcam eye-tracking data quality. Results will inform a more methodological understanding of webcam-based eye-tracking, clarifying whether its current limitations are intrinsic to the technology or can be mitigated through improved hardware and experimental control.

Keywords: Webcams, Eye-tracking, VWP, Lab, Competition, Spoken word recognition
Words: 3396

Online experimentation in the behavioral sciences has advanced considerably since its introduction at the Society for Computers in Psychology (SCiP) conference in Chicago, IL, in the mid-1990s (Reips, 2021), and its use has grown substantially in the years since. One methodological domain that has shown particular promise in moving online is eye tracking. Traditionally, eye-tracking studies required controlled laboratory settings equipped with specialized and costly hardware—a process that is both resource- and time-intensive. More recently, however, a growing body of research has shown that eye tracking can be successfully adapted to online environments (Bogdan et al., 2024; Bramlett & Wiener, 2024; Cruyssen et al., 2023; James et al., 2025; Prystauka et al., 2024; Slim et al., 2024; Slim & Hartsuiker, 2023; Vos et al., 2022; Yang & Krajbich, 2021; Özsoy et al., 2023). By leveraging standard webcams, researchers can now record eye movements remotely, making it possible to collect data from virtually any location at any time. This shift not only enhances scalability but also broadens access to more diverse and representative participant samples.

Webcam-based eye tracking is straightforward to implement for research purposes. It typically requires only a standard computing device—such as a laptop, desktop computer, tablet, or smartphone—equipped with a built-in or external webcam. Data collection is conducted through a web browser running specialized software that records and estimates eye movements in real time. The accessibility of webcam eye tracking has been greatly enhanced by its integration into several popular experimental platforms, including Gorilla (Anwyl-Irvine et al., 2019), PsychoPy/Psycho(Peirce et al., 2019), jsPsych (Leeuw, 2014) , PCIbex ((Zehr & Schwarz, 2022), and labvanced (Kaduk et al., 2023).

To reliably estimate where users are looking, webcam-based eye tracking typically relies on appearance-based methods, which infer gaze direction directly from visual features of the eye re-

*Send correspondence to: Jason Geller, drjasongeller@gmail.com.
Created with Quarto 1.7.29 and *preprint-typst*

gion (e.g., pupil and iris appearance). Recent work has extended these methods using deep learning to learn gaze–appearance mappings directly from data (Kaduk et al., 2023; Saxena et al., 2024). This contrasts with research-grade eye trackers, which use model-based algorithms combining infrared illumination with geometric modeling of the pupil and corneal reflections (Cheng et al., 2024).

The most widely used library for webcam eye tracking is WebGazer.js (Papoutsaki et al., 2016; Patterson et al., 2025). WebGazer.js is an open-source JavaScript library that performs real-time gaze estimation using standard webcams. It is an appearance-based method that leverages computer vision techniques to detect the face and eyes, extract image features, and map these features onto known screen coordinates during a brief calibration procedure. Once trained, gaze locations on the screen are estimated via ridge regression (Papoutsaki et al., 2016).

Although webcam eye-tracking is still relatively new, validation efforts are steadily accumulating and the results are encouraging. Researchers have successfully applied webcam-based methods to domains such as language (Bramlett & Wiener, 2025; Geller et al., 2025; Prystauka et al., 2024), judgment and decision-making (Yang & Krajbich, 2021), and memory (James et al., 2025) Overall, these studies replicate canonical effects and show strong convergence with findings from traditional lab-based eye-tracking systems.

However, there are a few shortcomings to web-based eye-tracking. First, effect sizes are often smaller than in lab-based studies (Bogdan et al., 2024; Degen et al., 2021; Kandel & Snedeker, 2024; Slim et al., 2024; Slim & Hartsuiker, 2022; Van der Cruyssen et al., 2024), which typically requires larger samples to achieve comparable power. Second, relative to research-grade eye-trackers, spatial and temporal precision are poorer: webcam approaches can commonly yield spatial accuracy of roughly 4° (Semmelmann & Weigelt, 2018) and temporal resolution can range from 50 ms-1000 ms (Geller et al., 2025; Semmelmann & Weigelt, 2018; Slim et al., 2024; Slim & Hartsuiker, 2023). These constraints make webcam eye-tracking less suitable for research that requires fine-grained spatial or temporal fidelity—for example, designs with many areas of interest (AOI) distributed across the screen or small AOIs [AOIs; James et al. (2025)], or tasks requiring precise moment-to-moment processing (Slim et al., 2024). Lastly, webcam eye-tracking yields high attrition rates. Looking at a number of webcam eye-tracking studies Patterson et al. (2025) found the attrition rate was on average 13%, with studies reporting a considerable range around this (Geller et al., 2025; Prystauka et al., 2024).

An open question is whether the limitations of web-based eye-tracking stem from the WebGazer.js algorithm itself or from environmental and hardware constraints—and, importantly, whether future improvements can address these issues. On the algorithm side of things, recent work (James et al., 2025) showed that modifying WebGazer.js insofar that the sampling rate is polled consistently and timestamps are aligned to acquisition (when data are received) rather than at completion (when processing finishes) markedly improves temporal resolution. Implementations of these changes in online experiment platforms (e.g., Gorilla and jsPsych) have brought webcam eye-tracking studies closer to the timing fidelity seen in lab-based eye-tracking. For instance, Prystauka et al. (2024), using the Gorilla platform, observed a 50 ms timing difference between lab-based effects and online effects while Geller et al. (2025) noted a 100 ms timing difference.

To our knowledge, no study has directly tested how environmental and hardware constraints impact webcam-based eye-tracking data. Slim & Hartsuiker (2023) provided some evidence suggesting that hardware quality may underlie some of these limitations, reporting a positive correlation between webcam sample rate and calibration accuracy. Similarly, Geller et al. (2025) found that participants who failed calibration more often reported using low-quality built-in webcams and working in suboptimal environments (e.g., natural lighting). Together, these findings suggest that both hardware and environmental factors may contribute to the increased noise commonly observed in online eye-tracking data.

Proposed Research

To address environmental and technical sources of noise in webcam eye-tracking, we plan to bring participants into the lab to complete a Gorilla-hosted webcam task under standardized conditions. We will manipulate two factors across two experiments. Experiment 1 will vary webcam quality (high- vs. low-quality external cameras). Experiment 2 varies head stabilization (with vs. without a chinrest). All sessions will be conducted under standardized conditions: identical ambient lighting, fixed viewing distance, the same display/computer model, and controlled network settings. This design allows us to isolate the causal effects of hardware and movement on data quality. Our key questions are whether higher-quality webcams and reduced head movement decrease noise, thereby (a) increasing effect sizes (higher proportion of looks), (b) yielding earlier onsets of established effects, and (c) reducing calibration failures/attrition rate. As noted above

To examine these factors, we replicate a paradigm widely used in psycholinguistics—the Visual World Paradigm [VWP; Cooper (1974); Tanenhaus et al. (1995)]. The VWP has been used quite successfully with webcam eye-tracking (Bramlett & Wiener, 2024; 2025; Geller et al., 2025; Prystauka et al., 2024). While there are slight variations in how the paradigm is implemented (Huettig et al., 2011), in the version most relevant to the current study, on each trial four images appear in one of four screen quadrants while a spoken word is played. Participants then select the picture that matched the utterance. Related to the current experiment, item sets are sometimes constructed so that the display contains a target (e.g., CARROT), a cohort competitor (e.g., CARRIAGE) , a rhyme competitor (e.g., PARROT), and an unrelated distractor (e.g., TADPOLE). A setup such as this allows one to examine how competition dynamics (i.e., CARRIAGE vs. TADPOLE or PARROT vs. TADPOLE) unfold over the time course of language processing (Colby & McMurray, 2023). Importantly, competition effects have previously been observed in webcam eye-tracking studies using a similar design (Geller et al., 2025) thus serve as good testing case herein.

Experiment 1

Slim & Hartsuiker (2023) and Geller et al. (2025) observed a relationship between webcam quality and calibration accuracy. Building on this, Experiment 1 will test how webcam quality influences competition effects in a single-word VWP/ Specifically, we ask whether a higher-quality webcam yields (a) a greater proportion of looks (i.e., stronger detectability of competition), (b) earlier emergence of the effect over time, and (c) lower attrition rates relative to a lower-quality webcam.

Hypotheses

We hypothesize several effects related to competition, onset, and attrition.

Competition Effects

(H1a) Participants will show a competition effect, with more looks directed toward cohort competitors than unrelated distractors. (H1b) Webcam quality (high vs. low) will influence the overall proportion of looks, with higher-quality webcams detecting a greater number of looks. (H1c) There will be an interaction between webcam quality and competition, such that the magnitude of the competition effect will be larger in the high-quality webcam condition than in the low-quality condition.

Onset Effects

(H2a) Looks to cohort competitors will emerge earlier than looks to unrelated distractors. (H2b) The onset of looks will occur earlier in the high-quality webcam condition than in the low-quality condition. (H2c) Consequently, the competition effect will emerge sooner in the high-quality webcam condition compared to the low-quality condition.

Attrition (H3) Attrition rates will be lower in the high-quality webcam condition than in the low-quality webcam condition.

Method

All stimuli (audio and images), code, and data will be stored on OSF (https://osf.io/cf6xr/overview).

Sampling Goal

We conducted an a priori power analysis via Monte Carlo simulation in R. Data from 21 participants, collected online using the Gorilla experimental platform during the development of the webgazeR package and employing the same stimuli and VVWP design, were used to seed the simulations. In these data, we observed a cohort effect of approximately 3%. Using this value as our seed, we collapsed the data across time bins to compute binomial counts per trial and fit a binomial generalized linear mixed model (GLMM) to obtain fixed-effect estimates. We then augmented the dataset by adding a between-subjects factor for webcam quality, with participants evenly assigned to high- and low-quality groups. In the high-quality webcam group, we modeled both a higher overall fixation rate and a larger cohort effect, whereas in the low-quality group the cohort effect was halved relative to the high-quality group. Simulated datasets were generated under this model, and the planned GLMM—including a condition_num × webcam_group interaction—was refit to each simulated dataset (5000). Power was estimated as the proportion of simulations in which the interaction term exceeded $|z| = 1.96$. The analysis script to run this power analysis is located here: https://osf.io/mywtq/overview. Results indicated that a total of 35 participants per group (N = 70) would provide approximately 90% power to detect the hypothesized reduction in the cohort effect and overall fixation rate under low-quality webcam conditions. We will therefore recruit participants until we have 35 in each group (N = 70 total). For the calibration analysis (see below), all participants who enter the study will be included.

Materials

VWP

Picture Stimuli. Stimuli were adapted from Colby & McMurray (2023) . Each set comprised four images: a target, an onset (cohort) competitor, a rhyme competitor, and an unrelated item (e.g., rocket, rocker, pocket, bubble). For the webcam study, we used 30 sets (15 monosyllabic, 15 bisyllabic).

Within each set, only the target and its onset competitor served as auditory targets once each, yielding two trial types: TCRU (target–cohort–rhyme–unrelated) and TCUU (target–cohort–unrelated–unrelated). This resulted in 60 trials total (30 sets × 2 targets per set). A MATLAB script generated a unique randomized list per participant, pseudo-randomizing display positions so that each image type was approximately equally likely to appear in any quadrant across subjects.

All 120 images were from a commercial clipart database that were selected by a small focus group of students and edited to have a cohesive style using a standard lab protocol (McMurray et al., 2010). Images were all scaled to 300 × 300 pixels.

Auditory Stimuli. Auditory stimuli were recorded by a female monolingual speaker of English in a sound-attenuated room sampled at 44.1 kHz. Auditory tokens were edited to reduce noise and remove clicks. They were then amplitude normalized to 70 dB SPL. . All .wav files were converted to .mp3 for online data collection.

Webcams. To manipulate recording quality, two webcams will be used:

- High-quality condition: A Logitech Brio webcam will record in 4K resolution (up to 4096 × 2160 px) with a 90° field of view, providing high-fidelity video suitable for accurate gaze estimation and pupil tracking.
- Standard-quality condition: A Logitech C270 HD webcam will record in 720 p resolution, producing video comparable to that of a typical laptop webcam, thereby simulating lower-quality online recordings.

Both webcams will be mounted in a fixed position above the monitor to maintain consistent framing across participants. Lighting will be standardized to ensure uniform image quality across all sessions.

Experimental Setup and Procedure

All tasks will be completed in a single session lasting approximately 30 minutes. The experiment will be programmed and administered in Gorilla (Anwyl-Irvine et al., 2019). Participants will be brought into a room in the Human Neuroscience Lab at Boston College and seated in front of a 23-inch Dell U2312HM monitor (1920 × 1080 px) approximately 65 cm from the screen. Audiotry information will be presented over Sony MDR-7506 headphones to ensure consistent audio presentation and minimize background noise. The experimental tasks will be fixed: informed consent, single word VWP, and a demographic questionnaire. The entire experiment can be viewed on Gorilla at this link:.

Before the main task, an instructional video will demonstrate the calibration procedure. Calibration will occur twice—once at the start and again after 30 trials—with up to three attempts allowed each time. In each calibration phase, participants will view nine calibration targets and five validation points, looking directly at each target as instructed. Participants will then complete four practice trials to familiarize themselves with the task. Each trial begins with a 500 ms central fixation cross, followed by a preview display of four images located in the screen's corners. After 1500 ms, a start button appears at the center; participants click it to confirm fixation before hearing the spoken word. The images remain visible throughout the trial, and participants indicate their response by clicking the image corresponding to the spoken target. A response deadline of 5 seconds will be used. Eye movements are recorded continuously during each trials. Following the main VWP task, participants will complete a brief demographic questionnaire, after which they will be thanked for their participation.

Data Preprocessing and Exclusions

We will follow guidelines outlined in Geller et al. (2025). At the participant level, individuals with overall task accuracy below 80% will be excluded. At the trial level, only correct-response trials (accuracy = 1) will be retained. Reaction times (RTs) outside ±2.5 SD of the participant-level distribution (computed within condition) will be discarded.

For eye-tracking preprocessing we will use the {webgazeR} package in R that contains helper functions to preprocess webcam eye-tracking data. All webcam eye-tracking files will be merged. Data quality will be screened via sampling-rate checks with very low-frequency recordings (e.g., < 5 Hz) by subject and trial excluded **(Bramlett & Wiener, 2024)**. We will quantify out-of-bounds (OOB) samples—gaze points outside the normalized screen (1,1)—and remove participants and trials with excessive OOB data ($> 30\%$). OOB samples will be discarded prior to analysis. In addition, Gorilla provides calibration/quality metrics ("convergence" and "confidence," both 0–1); trials with convergence < 0.5 or confidence > 0.5 will be excluded.

Areas of Interest (AOIs) will be defined in normalized coordinates as the four screen quadrants, and gaze samples will be assigned to AOIs. To create a uniform time base, data will be resampled into 100-ms bins. Trial time will be aligned to the actual stimulus onset by taking the audio onset provided by Gorilla. We then subtract 200 ms to approximate saccade programming and execution latency (Viviani, 1990), and an additional 100 ms due to silence prefixed to the audio recording.

For analysis, within each participant × trial × time bin we will compute, for each AOI, the number of valid gaze samples in that AOI ("successes") and the total number of valid samples in the bin ("trials"). These binomial counts (or their proportions) will serve as inputs to the statistical models and summaries; subject- and condition-level aggregates will be obtained by averaging across trials for descriptive plots.

Analysis Plan

Competition and Onset Effects

To analyze overall competition effects and onset latency, we will use generalized additive mixed models (GAMMs; Wood, 2017). GAMMs extends the generalized linear modeling framework by modeling effects that are expected to vary nonlinearly over time–a common feature in the VWP (Brown-Schmidt et al., 2025; Mitterer, 2025; Veríssimo & Lago, 2025). These models capture non-linear effects by fitting smoothing splines—or "wiggles"—to the data using data-driven, machine-learning-based methods. This approach reduces the risk of under and over-fitting and eliminates the need to prespecify polynomial terms, as required in traditional growth curve models (Mirman, n.d.). Importantly, GAMMs also allow researchers to account for autocorrelation in time-series data, which is especially critical in gaze analyses where successive samples are not independent. By modeling the autocorrelation structure, GAMMs provide more accurate estimates of temporal effects and prevent inflation of Type I error rates. In addition to this, fitting GAMMs allow us to estimate the onset of the competition effect in each condition (Veríssimo & Lago, 2025).

Gaze samples will be analyzed with a binomial (logistic) GAMM using the `bam()` function from the **{mgcv}** package (Wood, 2017). For visualization, we will employ functions from the {tidygam} package (Coretta, 2024), and use the **{onsets}** package (Veríssimo & Lago, 2025) to examine onset latencies. The dependent variable consisted of gaze counts to cohort and unrelated pictures, for each participant and in each 100 ms time bin. All analyses were conducted on a window ranging from −100 ms from target onset to 1200 ms.

We will fit a model including parametric terms for webcam type (effects-coded: high = 0.5, low = −0.5), item type (effects-coded: cohort = 0.5, unrelated = −0.5), and their interaction, capturing the overall (time-independent) effects. To examine how webcam type moderates the cohort effect over time, these two factors will also be combined into a single four-level factor.[1]Nonlinear, time-dependent effects will be modeled using factor smooths for time and for time-by-condition interactions, with condition treated as a categorical variable. To account for individual differences, we include random smooths by participant and random smooths for time by participant for each level of condition. This specification allows the model to capture (a) overall differences in fixation proportions between conditions (via the parametric terms), (b) dynamic, time-varying trajectories unique to each condition (via the smooth terms), and (c) participant-specific deviations from these group-level patterns (via the random smooths). While it is common to specify the maximal model (Barr et al., 2013), these can be costly when fitting GAMMS (Veríssimo & Lago, 2025).

To account for autocorrelation in the residuals, we will first fit the model without an autoregressive term in order to estimate the autocorrelation parameter (ρ). We will then re-fit the model including a first-order autoregressive process (AR(1)) to properly model temporal dependencies. Although using larger time bins can reduce autocorrelation, it does not eliminate it entirely, so explicitly modeling residual autocorrelation ensures valid statistical inference. Template code to fit the code is included below.

```
# quick rho estimate (fit once without AR to get residual ACF ~ lag1)

dat$cond4 <- interaction(dat$condition, dat$webcam, drop = TRUE)


m0   <- bam(cbind(fix, fail) ~ 1 + cond_c*cam_c +

       s(time, k = 10) +

              s(time, by = cond4 k = 10) +      # condition-specific curves
interaction

              s(participant, bs = "re") +        # random intercepts

              s(time, participant, by=cond4,  bs = "re"),  # random fucntional
smooths for time/subject by cond

            family = binomial(), method = "fREML",

            discrete = TRUE, data = dat, na.action = na.omit, select = TRUE)


rho  <- acf(residuals(m0, type = "pearson"), plot = FALSE)$acf[2]


# final model with AR(1) to handle within-series autocorrelation

m1 <- bam(cbind(fix, fail) ~ 1 + cond_c*cam_c +

              s(time, k = 10) +
```

[1]GAMs are inherently additive, meaning that interactions between nonlinear (smooth) terms cannot be estimated directly. To evaluate time-varying interactions or simple effects, it is therefore standard practice to combine relevant factors into a single composite factor and fit condition-specific smooths (Coretta & Casillas, 2024).

```
              s(time, by = cond4, k = 10) +      # condition-specific curves
interaction
              s(participant, bs = "re") +        # random intercepts
              s(time, participant, by=cond4,  bs = "re"),  # subject smooths
            family = binomial(), method = "fREML",
            discrete = TRUE, data = dat, na.action = na.omit, select = TRUE),
          discrete = TRUE, rho = rho, AR.start = dat$AR.start,
          data = dat, na.action = na.omit, select = TRUE)
summary(m1)
```

```
# Obain onsets in each condition (and their differences)
(onsets_comp <- get_onsets(model = m1,              # Fitted GAMM
                          time_var = "time",          # Name of time variabl
                       by_var = "conditionfactor",       # Name of condition/
group variable
                          difference = T,             # Obtain differences
between onsets
                          n_samples = 10000,          # Large number of
samples (less variable results)
                          seed = 1,                   # Random seed for
reproducibility
                          silent = T))                # Cleaner output in
documentation
```

Calibration

To examine whether webcam affects calibration rejection, we will fit a logistic regression model using the glm function and the code blow.

```
glm(calibration ~ webcam, family = binomial(link = "logit"))
```

Experiment 2

In Experiment 2, we will use the same low-quality webcam as in Experiment 1 and manipulate head stability by comparing a chin-rest condition to a no–chin-rest condition. Some online platforms [e.g., Labvanced; Finger et al. (2017)] mitigate head motion by warning participants when they move outside a predefined region; however, it remains unclear how such motion control interacts with WebGazer.js estimates of event detection and onset latency. We therefore test the following hypotheses regarding competition, onset, and attrition.

We hypothesize several effects concerning competition, onset, and attrition. Participants are expected to show a competition effect, with more looks directed toward cohort competitors than to unrelated distractors. The use of a chin rest is predicted to influence the overall proportion of looks. The competition effect is predicted to be larger when participants use a chin rest than when they do not. We also expect looks to cohort competitors to emerge earlier than looks to unrelated distractors, with overall gaze onsets occurring sooner in the chin-rest condition. Finally, we anticipate that attrition rates will be lower in the chin-rest condition compared to the no–chin-rest condition

Sampling goal, materials , procedure

The sampling goal, materials, and procedure are the same as Experiment 1. The difference is whether participants use a chin rest or not.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bogdan, P. C., Dolcos, S., Buetti, S., Lleras, A., & Dolcos, F. (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), 2213–2226. https://doi.org/10.3758/s13428-023-02143-z

Bramlett, A. A., & Wiener, S. (2024). The art of wrangling. *Linguistic Approaches to Bilingualism*. https://doi.org/https://doi.org/10.1075/lab.23071.bra

Bramlett, A. A., & Wiener, S. (2025). The Art of Wrangling: Working with Web-Based Visual World Paradigm Eye-Tracking Data in Language Research. *Linguistic Approaches to Bilingualism*, *15*(4), 538–570. https://doi.org/10.1075/lab.23071.bra

Brown-Schmidt, S., Cho, S.-J., Fenn, K. M., & Trude, A. M. (2025). Modeling spatio-temporal patterns in intensive binary time series eye-tracking data using Generalized Additive Mixed Models. *Brain Research*, *1854*, 149511. https://doi.org/10.1016/j.brainres.2025.149511

Cheng, Y., Wang, H., Bao, Y., & Lu, F. (2024, April). *Appearance-Based Gaze Estimation With Deep Learning: A Review and Benchmark* (Issue arXiv:2104.12668). https://doi.org/10.48550/arXiv.2104.12668

Colby, S. E., & McMurray, B. (2023). Efficiency of spoken word recognition slows across the adult lifespan. *Cognition*, *240*, 105588. https://doi.org/10.1016/j.cognition.2023.105588

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Coretta, S. (2024). *tidygam: Tidy Prediction and Plotting of Generalised Additive Models*. https://github.com/stefanocoretta/tidygam

Coretta, S., & Casillas, J. V. (2024). A tutorial on generalised additive mixed effects models for bilingualism research. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.23076.cor

Cruyssen, I. Van der, Ben-Shakhar, G., Pertzov, Y., Guy, N., Cabooter, Q., Gunschera, L. J., & Verschuere, B. (2023). 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*. https://doi.org/10.3758/s13428-023-02221-2

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017,). LabVanced: A Unified JavaScript Framework for Online Studies. *International Conference on Computational Social Science (Ic2s2)*.

Geller, J., Prystauka, Y., Colby, S. E., & Drouin, J. R. (2025). Language without Borders: A Step-by-Step Guide to Analyzing Webcam Eye-Tracking Data for L2 Research. *Research Methods in Applied Linguistics*, *4*(3), 100226. https://doi.org/10.1016/j.rmal.2025.100226

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

James, A. N., Ryskin, R., Hartshorne, J. K., Backs, H., Bala, N., Barcenas-Meade, L., Bhattarai, S., Charles, T., Copoulos, G., Coss, C., Eisert, A., Furuhashi, E., Ginell, K., Guttman-McCabe, A., Harrison, E. (., Hoban, L., Hwang, W. A., Iannetta, C., Koenig, K. M., … de Leeuw, J. R. (2025). What Paradigms Can Webcam Eye-Tracking Be Used For? Attempted Replications of Five Cognitive Science Experiments. *Collabra: Psychology*, *11*(1), 140755. https://doi.org/10.1525/collabra.140755

Kaduk, T., Goeke, C., Finger, H., & König, P. (2023). 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

Leeuw, J. R. de. (2014). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

Mirman, D. *Growth curve analysis and visualization using R*.

Mitterer, H. (2025). A web-based mouse-tracking task for early perceptual language processing. *Behavior Research Methods*, *57*(11). https://doi.org/10.3758/s13428-025-02827-8

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016, July). WebGazer: Scalable Webcam Eye Tracking Using User Interactions. *International Joint Conference on Artificial Intelligence*.

Patterson, A. S., Nicklin, C., & Vitta, J. P. (2025). Methodological recommendations for webcam-based eye tracking: A scoping review. *Research Methods in Applied Linguistics*, *4*(3), 100244. https://doi.org/10.1016/j.rmal.2025.100244

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Prystauka, Y., Altmann, G. T. M., & Rothman, J. (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), 3504–3522. https://doi.org/10.3758/s13428-023-02176-4

Reips, U.-D. (2021). Web-based research in psychology: A review. *Zeitschrift Für Psychologie*, *229*(4), 198–213. https://doi.org/10.1027/2151-2604/a000475

Saxena, S., Fink, L. K., & Lange, E. B. (2024). Deep Learning Models for Webcam Eye Tracking in Online Experiments. *Behavior Research Methods*, *56*(4), 3487–3503. https://doi.org/10.3758/s13428-023-02190-6

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451–465. https://doi.org/10.3758/s13428-017-0913-7

Slim, M. S., & Hartsuiker, R. J. (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), 3786–3804. https://doi.org/10.3758/s13428-022-01989-z

Slim, M. S., Kandel, M., Yacovone, A., & Snedeker, J. (2024). Webcams as Windows to the Mind? A Direct Comparison Between In-Lab and Web-Based Eye-Tracking Methods. *Open Mind*, *8*, 1369–1424. https://doi.org/10.1162/opmi_a_00171

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science (New York, N.Y.)*, *268*(5217), 1632–1634. http://www.ncbi.nlm.nih.gov/pubmed/7777863

Veríssimo, J., & Lago, S. (2025, April 29). *A novel method for detecting the onset of experimental effects in visual world eye-tracking*. PsyArXiv. https://doi.org/10.31234/osf.io/yk4xb_v3

Viviani, P. (1990). Eye movements in visual search: cognitive, perceptual and motor control aspects. *Reviews of Oculomotor Research*, *4*, 353–393.

Vos, M., Minor, S., & Ramchand, G. C. (2022). Comparing infrared and webcam eye tracking in the Visual World Paradigm. *Glossa Psycholinguistics*, *1*(1). https://doi.org/10.5070/G6011131

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman, Hall/CRC.

Yang, X., & Krajbich, I. (2021). Webcam-Based Online Eye-Tracking for Behavioral Research. *Judgment and Decision Making*, *16*(6), 1485–1505. https://doi.org/10.1017/S1930297500008512

Zehr, J., & Schwarz, F. (2022). PennController for Internet Based Experiments (IBEX). *Open Science Framework*. https://doi.org/10.17605/OSF.IO/MD832

Özsoy, O., Çiçek, B., Özal, Z., Gagarina, N., & Sekerina, I. A. (2023). Turkish-German heritage speakers' predictive use of case: webcam-based vs. in-lab eye-tracking. *Frontiers in Psychology*, *14*, 1155585. https://doi.org/10.3389/fpsyg.2023.1155585