

Methodological Protocol for High-Quality Data Collection during Eye Tracking in Human Reading

Cengiz Acartürk¹, Duygun Erol Barkana², Funda Yildirim³, Simone Borsci³, Eda Nur Capkan^{3,4}, Kristina Cergol⁵, Ramunė Kasperė⁶, Dorota Klimek-Jankowska⁷, Oksana Ivchenko⁸, Yue Li³, Paula Luegi⁹, Jurgita Motiejūnienė⁶, Jamal Nasir¹⁰, Marijan Palmović¹¹, Oleksii Turuta¹²

¹Centre for Cognitive Science, Jagiellonian University ²Department of Electrical and Electronics Engineering, Yeditepe University ³Section of Cognition, Data and Education, University of Twente ⁴Department of Psychology, Istanbul Ticaret University ⁵Faculty of Teacher Education & Laboratory for Psycholinguistic Research, University of Zagreb ⁶Faculty of Social Sciences, Arts and Humanities, Kaunas University of Technology ⁷Institute of English Studies, University of Wrocław ⁸CNRS, University of Lille ⁹Center of Linguistics, School of Arts and Humanities, University of Lisbon ¹⁰School of Computer Science, College of Science and Engineering, University of Galway ¹¹Laboratory for Psycholinguistic Research, Department of Speech and Language Pathology, Faculty of Education and Rehabilitation Sciences, University of Zagreb ¹²Institution of Computer Science and AI, V. N. Karazin Kharkiv National University

✉ Corresponding Author : Cengiz Acartürk <cengiz.acarturk@uj.edu.pl>

Citation

Acartürk, C., Barkana, D.E., Yildirim, F., Borsci, S., Capkan, E.N., Cergol, K., Kasperė, R., Klimek-Jankowska, D., Ivchenko, O., Li, Y., Luegi, P., Motiejūnienė, J., Nasir, J., Palmović, M., Turuta, O. Methodological Protocol for High-Quality Data Collection during Eye Tracking in Human Reading. *J. Vis. Exp.* (235), e70937, doi:10.3791/70937. (2026).

Abstract

Eye tracking is a central methodological approach for studying attention and language processing during reading. The validity and reproducibility of results depend on data quality, clear methodology, and consistent preprocessing. This article presents a comprehensive, step-by-step protocol for collecting high-quality eye-tracking data in reading studies. The protocol begins with laboratory environment preparation and equipment setup, then proceeds to selecting and configuring eye-tracking hardware and software, and to stimulus preparation and experimental design. Participant management, informed consent, and documentation of linguistic and psychometric characteristics are conducted. Calibration and validation procedures are then performed, with participant comfort and data quality closely monitored online during data collection. The preprocessing pipeline includes exporting raw data in standard formats, data cleaning, fixation, and saccade detection (employing velocity-based and dispersion-based algorithms) with clearly defined parameters, visual inspection for signal loss and noise control, vertical drift correction for multiline text, and extraction of standard reading measures (fixation duration, saccade metrics, word skipping, and regression rates). Throughout the protocol, alignment between research questions and hardware specifications, management of drift through regular checks, and detailed reporting of all critical parameters and decisions are particularly emphasized. Successful data collection typically results in stable validation accuracy within set limits, low data loss rates, and interpretable distributions of fixation durations and saccade behavior that reproduce well-known effects related to word length, frequency, and predictability. The protocol is compatible with desktop eye trackers, common software environments, and both head-stabilized and head-free arrangements, addressing practical issues such as lighting conditions, reflections from glasses, and changes in participant posture. This study presents a clear, practical protocol that consolidates methodological guidance into a unified workflow. It aims to support reliable eye-tracking data collection in reading research, promote

replicability and meta-analyses, improve consistency across laboratories and groups, and ultimately strengthen the scientific credibility of reading research.

Introduction

Eye tracking is a central method for studying attention and language processing during reading, as fixations and saccades indicate the moment-to-moment allocation of visual and cognitive resources to the stimuli. Classic and contemporary reviews document robust links between eye movements and theoretical constructs such as the perceptual span, preview benefit, and models of eye-movement control^{1,2}. To conduct reliable studies examining these links, methodological rigor in equipment choice, experimental design, calibration and validation, and preprocessing is essential; otherwise, data quality degradation biases estimates and undermines replicability³.

Over the last decade, several primers and best-practice papers have consolidated the field's methodological knowledge. A recent fundamentals series connects theory to measures, discusses operationalization and algorithmic variability, and provides guidance for choosing eye trackers for specific research questions^{4,5,6,7,8}. Complementary guidelines present best practices for selecting hardware, ensuring accuracy and precision, defining measures, and reporting, with recommendations for reducing threats to validity⁹. Introductory resources for reading research also offer accessible overviews of experiment design, common measures, and analysis pipelines^{10,11}. In applied linguistics and reading, domain-specific reviews summarize the selection of measures and trackers and highlight task-dependent constraints^{12,13,14,15,16,17}.

This article translates that guidance into a stepwise protocol for high-precision eye-tracking data collection in reading studies. Unlike existing resources that primarily offer broad, theoretical best-practice guidelines or focus narrowly on specific analytical techniques, this protocol operationalizes these concepts into an end-to-end, actionable workflow. It aims to facilitate daily laboratory practice by providing concrete, sequential instructions. In terms of applicability, this protocol is specifically optimized for highly controlled laboratory reading experiments utilizing high-frequency, desktop-based eye trackers (e.g., 500-2000 Hz) with head-stabilized setups (chin or forehead rests). It is best suited for research demanding strict data-quality constraints, high spatial accuracy, and precision—such as studies analyzing precise word-level fixations or fine-grained saccade metrics.

The use of head-free or naturalistic setups has also been gaining popularity, including in reading research, while we limit the scope of the article to laboratory studies. We cover hardware and software selection, lab setup, stimulus preparation, participant management, calibration and validation, online quality control, and

preprocessing (including event detection and drift correction). By following the protocol, researchers can document and achieve data quality consistent with current standards, thus improving transparency, credibility, and reproducibility of reading-related findings.

Protocol

Ethics permission for the COST CA21131 data collection was granted by the Research Ethics Committee at the Faculty of Philosophy, Jagiellonian University, Poland, with the protocol number 221.0042.18_2024. Informed consent was obtained from all participants.

1. Preparing the Lab environment and setting up equipment

1. Preparing the lab environment
 1. Check the light and environmental noise conditions. Windows in the lab environment may cause sudden pupil dilation changes and light reflections on the eye tracker camera and the monitor. If possible, gather data in a windowless room or use blackout curtains to block sunlight. Use artificial lighting and eliminate reflective light sources. Verify that no reflections appear on the monitor; adjust blinds/angles until none remain. Ensure the room is quiet and prevent interruptions (e.g., signage, door locks). Use external signals to avoid interruptions (e.g., put a signal on the outside with "Experiment ongoing. Please do not knock on the door or enter the room.")
 2. Check physical ergonomics. Use a stable, non-swiveling chair; adjust table height (if adjustment is available) to maintain posture and avoid calibration loss from chair rotation.
 3. Avoid social presence. The presence of an experimenter in the same room may distract the participant. Separate the control room from the participant room, or keep it out of view, to minimize social presence effects.
 4. Check air ventilation. Carry out airing before and after experiment sessions. Check the room temperature, since participants will need to remain still for some time.

2. Selecting eye-tracking equipment

1. Check the compatibility between research goals and measures. The goals of a specific study, thus the target

measures (dependent variables), are expected to align with the capabilities of the eye-tracking equipment. The institutional availability of the equipment might introduce limitations to the scope of the intended study. Check the three major component characteristics (sampling rate, precision, and accuracy) of the eye-tracking system and their potential for fulfilling the research needs. The main point is to ensure that the capability of the eye-tracking equipment is compatible with the goals of the study.

NOTE: Sampling rate = The sampling rate is expressed in Hz and reflects the size of the samples collected by the eye tracker during a period: a 60 Hz eye tracker collects samples every 16 ms, while a 500 Hz eye tracker collects samples every 1 ms. This affects the different types of measures, since low-frequency eye trackers might include saccades that started during a 16-ms period as fixations, or vice versa. Sampling is sometimes conflated with data accuracy, but these are distinct properties.

Precision = The consistency with which the eye tracker is able to collect data over time.

Accuracy = The amount of error between the point of gaze detected by the eye tracker and the actual point of gaze.

As an example, in studying fixation landing sites on words, which are described in terms of the position of gaze on the letters of a word, requires a much higher sampling rate (≥ 500 –2000 Hz), precision, and accuracy than a study solely focusing on the percentage of regressive saccades (compared to progressive saccades) during reading.

2. Determine the methodology of eye-movement recording. Use a desktop eye tracker with a high sampling rate, precision, and accuracy. Use head stabilization with a chinrest or a forehead rest. In case of using head-free eye tracking (remote eye tracking), use stickers, if available, on the participant's forehead to improve face tracking; also, avoid reflective stickers unless recommended by the tracker vendor.

NOTE: The use of wearable eye-tracking equipment has been rare in reading studies due to its relatively low sampling rate, precision, and accuracy compared to desktop eye trackers, as well as challenges in analyzing eye-movement data. However, head-free or naturalistic setups have been gaining popularity in the past decade, accompanied by improved data quality and automated analysis pipelines, thus extending the scope of studies on eye movements during reading toward naturalistic settings and reading in VR environments.

3. Setup of hardware and software

1. Setup hardware. Measure the distances between the participant's face, monitor, and eye tracker. Set viewing distance per manufacturer (e.g., ~60–70 cm) and required configuration geometry, preferably using a flat 21–27 inch LCD monitor (e.g., 24 inch at 60 cm, ~48° x 28° field of view) so that all text remains within the central visual field and the tracker's effective range. Check monitor tilt and rotation, and the alignment between the gaze vector and its corresponding location on the screen. Set the monitor size and refresh rate to the target values. Set the screen resolution to the specified value. Measure monitor luminance on a white background at the specified distance. Record the luminance value in the session report.
1. Where needed, prefer LCD monitors with high refresh rates (120 Hz or higher, e.g., 144–240 Hz), especially for gaze-contingent tasks and verified timing. Disable variable refresh technologies (G-Sync/FreeSync). If the eye tracker is used by different researchers for different experiments, it is important to keep track of any changes made during each data collection. At the end of each data collection, the setup should be returned to its original state at the beginning of the session, or as defined by the lab protocol.
2. Setup software. Most eye-tracker manufacturers provide relevant software, Software Development Kits (SDKs), and/or libraries for compatibility and integration with open-source software.

NOTE: Recently, specific setups have remained a technical challenge requiring custom solutions. For instance, time synchronization of an eye tracker with complementary methods of measurement, such as EEG and fNIRS, requires special hardware and/or software solutions to ensure an appropriate level of time-locking between the eye-tracker clock and the stimulus-presentation clock, and/or screen refresh-desynchronization. Therefore, most manufacturer tools provide an environment for designing experiments that meets researchers' basic needs, such as presenting the stimuli and recording eye movements. A list of currently available software packages is presented in the **Table of Materials**.

2. Designing and piloting the experiment

1. Preparation of stimuli
 1. Prepare reading materials. Reading materials are usually presented in the form of sentences or paragraphs,

depending on the goals of the study. Consider excerpting reading materials from available sources rather than manually creating text, as long as this method aligns with the goal of the study.

2. Check licensing conditions for text re-use. Check the licensing conditions for all texts. If the materials from published experiments are open source and thus publicly available to anyone who wants to use them, then permission is not necessary. Otherwise, obtain the required permissions before use. It is important to check whether stimuli can be used for the experiment and whether they can be shared as supplementary materials.
3. Prepare formatted stimuli. Formatting text stimuli requires using appropriate font properties (type, size, letter, word, and line spacing). Use a monospace font with a proper font size and spacing, considering the settings used in the literature.

NOTE: Monospace fonts, such as Courier New, are recommended fonts in reading experiments that use word-based Areas of Interest (AOIs), as these fonts provide the same pixel size for different alphanumeric characters and symbols within the Latin alphabet. The researcher should check the compatibility of the font, the interface, and the specific properties of the stimuli, such as the presentation of the stimuli in non-Latin characters or right-to-left reading directions, and whether the experiment software tools provide a methodology for performing automatic IA (Interest Area) or AOI (Area of Interest) specification.

2. Design experiment procedure

1. Preparation of instructions

NOTE: Instructions are important as they have a significant impact on eye movements during reading. The use of task instructions, even in minimal form, is crucial. A lack of explicit instructions may cause unintended consequences that affect participants' gaze behavior. For instance, one participant might have a latent assumption about careful reading, while the other may perform skimming rather than natural reading for comprehension.

1. Present standardized on-screen instructions and log instruction onset time and duration; avoid verbal prompts during recording. A common practice is to ask the participant to read for understanding. In case the participants are expected to read as they do in daily settings, inform the participants that they are expected to read naturally; that is, they do not need to avoid blinking or rereading. Also, inform participants if there is any additional task, such

as comprehension questions (which are important to guarantee that participants are reading for comprehension). Instructions may vary not only by task but also by the phase of the experiment (e.g., training vs. main task). To illustrate this, an example of instructions used during a training phase is provided below¹⁸.

We will now begin with a training session. For this, you will need the four marked keys (A, B, C, D). Further instructions will be shown to you shortly. After the training, your performance will be evaluated, which will take one to two minutes. If your training performance is not sufficient, the last part of the training will be repeated. If it is sufficient, I will return to the room. When you are ready, simply press the space bar.

2. It is up to the researcher to use such instructions for the training sessions of experiments that are designed for studying reading. Nevertheless, use the instructions for the experiment session. An example of experiment-phase instructions used in a general relevance instruction condition (both with and without a question in the text) may look like the one below¹⁹.

You will read a set of short expository texts. We want you to read the text carefully, understanding as much of the text as possible. Later, after reading, you will be asked to give an oral summary about the main ideas of the text to see how well you understood what you have read.

Another example instruction is given below for an experiment that employs sentence-reading tasks²⁰.

Welcome to our reading study. In this study, you will be presented with approximately a hundred sentences. They were selected randomly from various sources. They will be presented in random order. Read them silently, for understanding the meaning, at your own, natural pace of reading, in a similar way you would read sentences in a book. Following some of the sentences, you may be asked comprehension questions. Please reply to comprehension questions. If you do not know the answer, skip the question. There will be a break in the middle of the session. You may ask for a break or leave the session at any time if you feel uncomfortable. If you have any questions, please inform the experimenter. Otherwise, press a button to start the procedure.

2. Design trials and blocks. Design the flow of the experiment by considering the presentation of trials in blocks and randomization of the order of trials within and across the blocks. Use a fixation cross before each trial (500-800 ms) to ensure that the participants' initial fixations are on the same location on the display and away from any part of the text.

The design of an eye-tracking experiment, in which pupil size is a dependent variable, requires adjusting the flow of the experiment accordingly, as pupil size is not a momentary response like a fixation-saccade sequence. Specifically, in the case of the use of pupil size as a dependent variable, it is necessary to introduce a pre-onset baseline window and black screen intervals with appropriate durations to ensure stabilization of pupil size before the presentation of the stimuli and the subsequent trial. The researcher should check recent studies on pupillometry guidance to ensure that the recorded pupil data can be appropriately interpreted as a dependent variable^{21,22}.

3. Design breaks between blocks. Introduce drift-correction screens and optional short breaks between experiment blocks to ensure data quality and reduce fatigue. Insert drift checks after a prespecified number of trials, or when accuracy visibly degrades (check if there are short keys to force calibration/validation during the experiment); re-validate if drift $>1^\circ$ of visual angle (mean of x and y coordinates).
4. Identify target independent variables. In reading studies, specific properties of text, such as word frequency and the sentential predictability of words, comprise the main target independent variables. List the independent variables of the experiment, depending on the goals of the study. In statistical analyses, independent-variable metrics may require value normalization. Assess the relevance of each variable based on the properties of the stimuli.
 1. Exclude words at the beginning of the text line from the analysis, since fixations on those words are usually not accurate: they may reflect the first tentative fixation in the text/sentence or the first fixation after a line break (what is called a return-sweep), which are not, in any case, accurate and are usually followed by corrective fixations. Exclude the last word from each line because those words are subject to less visual crowding (from subsequent words) and because fixations on end-of-line words often reflect the termination of a trial or the programming of a return sweep to the next line of text.

NOTE: First fixation duration might be less significant in languages with many function words, such as articles before nouns (the analysis region should include the article and the noun, but the first fixation will be less important, as it may reflect only the processing of the article).

5. Identify target dependent variables. In reading studies, a set of eye-movement metrics, such as the first fixation duration, single fixation duration, gaze duration, regression percentage, and word skipping probability, comprises the main target-dependent variables in connection to the goals of the study. Specify the eye-movement metrics to be analyzed. Use these metrics as the backbone of the statistical analysis.

NOTE: Avoid using any task that requires the participant to type on the keyboard, as it may require leaving the chinrest, thus losing the calibration. Instead, participants may use the mouse to select their response. In cases where oral feedback from the participant is necessary during eye-movement recording, use a forehead rest instead of a chinrest, and ask the participant to respond verbally or allow single-key responses while remaining on the forehead rest.

3. Implementation of the experiment procedure

1. Select stimulus presentation software. Set up the experiment using appropriate stimulus presentation software. Both manufacturer-provided tools and general-purpose experiment design tools can be used to present text and record eye movements. When choosing software, consider the available hardware integrations (e.g., native support for their eye tracker), the degree of timing accuracy required, and their preferred level of scripting versus GUI-based design. Regardless of the platform, the software must support reliable synchronization between stimulus events (e.g., word onset, sentence onset) and the eye-tracking data stream so that fixations and saccades can be aligned with the corresponding textual elements during analysis. If using external software, check the compatibility between the eye-tracker software and the external software, especially for data visualization and data exportation.

NOTE: Examples of software: Experiment Builder (SR Research, for EyeLink systems) provides a graphical interface with built-in support for drift checks, calibration routines, and event messages sent to the recording computer. Open-source platforms such as PsychoPy and OpenSesame offer similar functionality and are hardware-agnostic, as they allow designing trials via

a graphical interface or scripting (e.g., Python in the case of PsychoPy), control stimulus timing, and define regions of interest, aka Interest Areas IAs or Areas of Interest AOI, around words or lines of text. Sometimes, the data visualization and data exportation linkage is not straightforward for researchers at the beginning of their careers, and, therefore, it may be easier to use the manufacturer's software instead. Consequently, an important consideration in choosing eye trackers is the level of support provided by the different manufacturers or the community of developers.

2. Pilot the experiment. Test the implemented experiment procedure and check the recorded data to ensure that the experiment runs as expected. Run pilot studies in the lab environment to ensure that the experiment runs as expected by monitoring predefined success criteria. Check the recorded data to confirm that everything needed for the study is being registered.

3. Recruiting and informing participants

1. Selection of participants

1. Determine vision condition requirements. Recruit participants with normal or corrected-to-normal vision. It is possible to accommodate glasses and contact lenses, which have minimal reflections (anti-glare, monitor angle) per manufacturer guidance. Document correction type (glasses, contact lenses) for covariate checks.

NOTE: Usually, rigid contact lenses (which are no longer very common) are very hard to track. Although contact lenses are not a problem, sometimes a bubble of air can form between the lens and the cornea, causing reflections that disrupt tracking. Glasses with shiny frames may also interfere with tracking. For shiny glass frames, reduce the field-of-view detection in the eye-tracker software.

2. Screen participants based on specific individual traits, such as age, reading proficiency, language background, handedness, and common exclusion and inclusion criteria, such as dyslexia, bilingualism, or reading disorders, may influence eye-movement data. Therefore, depending on the research goals, specify the conditions under which a reader is an appropriate participant for the experiment, thus the inclusion and exclusion criteria (see Discussion).
3. Check visual impairment conditions. Ensure that the participant does not have any uncorrectable visual impairment, such as cataracts (even after surgery), glaucoma, or macular degeneration.

4. Check neurological conditions. Ensure that the participant does not have any condition that makes them uncomfortable with eye-movement recordings. Most eye trackers use near-IR LEDs/lasers or certified NIR LED/VCSEL illumination, which are considered eye-safe. Confirm the device's eye-safety certification (e.g., Class 1 IEC 60825-1, IEC/EN 62471, or similar). Inform the participant about the equipment's safety regulations (such as the availability of a CE certificate from the manufacturer, commonly known as an EU Declaration of Conformity, which is a document issued by the manufacturer to certify that a product complies with all necessary European health, safety, and environmental regulations).

2. Establishing hygiene

1. Take precautions related to personal distance. If needed, wear disposable gloves and a mask, as running experiment protocols may require reduced personal distance.
2. Disinfect equipment. Disinfect the chinrest and other accessories using alcohol wipes as per current institutional guidance. Be careful not to use chemicals that may harm the equipment. Check the manufacturer's specifications. Participants may feel more comfortable if cleaning occurs in their presence, specifically for the chinrest and headrest, the mouse, or any other part of the equipment they will be in contact with.

3. Informing the participant

1. Inform the participant about the lab environment. Design a verbal instruction session in which the experimenter introduces the lab environment and the eye-tracking equipment to the participant.
2. Inform the participant about the experiment. Present a clear explanation of the experiment context (the goal and approximate duration of the study; this should be checked during the pilots), the incentive or payment procedure, the contribution of the participant to the experiment, and their rights to follow up on the results of the study in the long run. Inform the participant about dos and don'ts during a recording session. Ask the participant to try not to move their head too much during the recording (in case of no head stabilizer) or not to leave the chinrest, stating that these precautions are for high-quality recording of the data (otherwise, the participants can always stop the experiment and quit the session without providing a reason).
3. Obtain informed consent. Give participants the ethics approval and consent documents and explain them

before the session begins. For adults, obtain written informed consent after ensuring they understand the study's purpose, what will happen, and their rights, including the option to volunteer, data privacy, and the right to withdraw. For minors participating, obtain written consent from a parent or guardian and the minor's agreement. Continue with calibration and recording only after all required consent and assent forms are completed. Informed consent procedures vary considerably between adult and minor participants. Therefore, the researcher should follow local institutional guidelines.

NOTE: If a pause or break occurs during the experiment, revalidation of the calibration is required; if revalidation fails, recalibration is required.

4. Performing Calibration and Validation

1. Positioning the participant and adjusting the equipment

1. Position participant. Ask the participant to sit comfortably. Adjust the chair/table and/or the chinrest to maintain a steady head position. Inform the participant that it is the expected head position during calibration, validation, and recording phases.
2. Adjust equipment. Adjust the eye-tracking equipment by aligning the eye tracker camera so the participant's pupils are clearly visible and detectable. Some eye-tracking systems perform this alignment automatically, while others may require manually adjusting the focus of the camera.

NOTE: The experimenter should practice equipment adjustment and calibration before starting real data collection to avoid burdening the participant with these steps.

2. Running calibration

1. Inform participant. Inform the participant about the calibration procedure (e.g., stating that a series of target markers will be displayed, and that the participant is required to follow them for approximately a minute or so).

NOTE: It is recommended to use caution with the language while making the adjustments. For instance, the experimenter emphasizes that the procedure is for calibrating the eye tracker for the participant rather than calibrating the participant for the eye tracker. Moreover, it is important to ensure that no video/image is recorded with the eye-tracking equipment. Although the participant's eye is visible, only the eye's position, converted into coordinates, will be recorded.

2. Run calibration procedure. Instruct the participant to track the target markers and to keep their head position constant; for instance, not to lean backward after the end of the calibration procedure. Instruct participants to fixate on the center of the calibration symbol and not to move their eyes until the symbol moves. Participants tend to anticipate the symbol's movement, especially during validation, which may increase errors and lead to unnecessary calibration steps. Include this explanation in the instructions or in the explanation of the calibration/validation process.

NOTE: Manufacturers offer different types of symbols for the calibration process (dots, child-friendly symbols, static or shrinking symbols, etc.) and different calibration setups (i.e., the number of calibration points). It is important to check calibration differences with these different calibration setups and choose the one that best fits the needs of a specific study. For instance, if the stimuli use the entire screen (e.g., a text with many paragraphs), it is important to have a calibration setup that ensures good calibration from top to bottom, including the corners. However, if stimuli are sentences presented at the center of the screen, a calibration setup with fewer points may suffice. Additionally, calibration accuracy for the horizontal, or x-coordinate, is more important for the latter than for the former, which requires good accuracy in both vertical and horizontal coordinates.

3. End calibration procedure. Inform the participant about the validation procedure after completing the calibration procedure.

NOTE: In case of persistent failure in calibration, check the eye tracker's camera alignment, lighting conditions, and participant positioning. For adult participants, reflective makeup, glasses, or contact lenses may interfere with tracking. Reassure the participant, allow short breaks, or repeat the calibration as needed. If problems persist, consider restarting the eye-tracking system. If problems persist after restarting the eye-tracking system and no reason was found for the calibration failure after verifying the configuration parameters, inform the participant that the eye-tracking technology has not been developed sufficiently, and that it is an issue related to the status of technology, not the participant.

3. Performing validation

1. Run validation procedure. Run the validation procedure immediately after calibration.

2. Check errors. Check the average gaze position error and ensure that it is within acceptable limits. Typically, a visual angle of less than 1 degree ($<1^\circ$) is desired, which corresponds to approximately 4 characters (meaning that 0.5 corresponds to ~ 2 characters). Perform recalibration if the errors are outside these limits. For instance, accept calibration if the mean error is $\leq 0.5\text{--}1.0^\circ$ and the maximum/per-point error is $\leq 1.5\text{--}2.0^\circ$; otherwise, recalibrate. Specify the re-validation cadence (e.g., after each block or interruption). To exemplify criteria for rating the quality of a calibration by a manufacturer, SR Research, errors are considered acceptable for most tasks in case of an average error $<1^\circ$ and a maximum error $<1.5^\circ$; fair in case of an average error $<1^\circ\text{--}1.5^\circ$ and a maximum error $<1.5^\circ\text{--}2^\circ$; and poor when greater than these values.
3. Inform participant. Instruct the participant to keep their position as similar as possible and inform them about the beginning of the recording session.

5. Data collection

1. Running data collection procedure
 1. Start recording session. Start the recording session and avoid disrupting the participant while they are performing the task, such as providing verbal instructions.
 2. Monitor participant. Monitor participant comfort during data recording. Ensure that the participant maintains a stable posture without strain or fatigue. Watch for signs of distraction, eye dryness, or discomfort from the chinrest/forehead rest.
 3. Monitor data recording. Monitor data quality during the recording. Recalibrate if the average validation error or online accuracy exceeds 1° (or a prespecified accuracy threshold), or if the participant shifts posture.
 4. Keep a log. Keep a log of the recording session by taking notes of any anomalies observed during the experiment (e.g., an unusually long calibration procedure or an unintended interruption during the recording) for post-experiment inspection, using a session log schema to monitor the time of events, interruptions, the number of recalibrations performed and their errors, drift checks, and lighting changes.
2. End of the session
 1. Save data. Save raw gaze data at the end of the experiment session, preferably using a systematic method to label the session (e.g., the date of the experiment and the participant number, for instance

25071801 or 250718-P01 for participant 01 on 18 July 2025; avoid using personally identifiable information about the participant).

2. Farewell. Thank the participant and address any questions they may have about the experiment; sanitize equipment (see step 3.2.2).
3. Check data. Check the data and the logs to ensure the session was completed correctly.

6. Data preprocessing and quality check

1. Exporting, organizing, and cleaning of raw data
 1. Export data. Export raw gaze data in a standard format (e.g., .csv, .tsv, .edf) and standardize your session ID (e.g., 25071801 or 250718-P01, see 5.2.1).
 2. Clean data. Clean raw data by applying algorithms for blink detection and missing data detection, using built-in tools provided by the manufacturer's software or using custom scripts, adjusted to the device or sampling rate.
- NOTE:** Create backups of raw data while running the cleaning routines.

2. Detecting eye-movement events
 1. Select eye-movement event detection algorithm. Select and run the procedures for fixation detection, adjusting the algorithms' parameters as needed. The representative algorithms for fixation identification mainly involve velocity-based algorithms (e.g., I-VT Velocity-Threshold Identification) and dispersion-based algorithms (e.g., I-DT Dispersion-Threshold Identification), among others²³. Each algorithm has its own parameters.

NOTE: Manufacturers' software usually assumes a specific algorithm and a specific set of commonly used parameters by default. It is the researcher's responsibility to select an appropriate algorithm and its parameters and report it in the description of the methodology of the study, such as dispersion/velocity thresholds, a minimum fixation duration threshold (e.g., 60 ms or 100 ms) to eliminate fixations shorter than this value, a maximum fixation duration threshold (e.g., 1000 ms) to eliminate fixations longer than this value, the smoothing window used, and the merging rules for adjacent fixations²³.

2. Identify illegitimate fixations. Remove or label the fixations outside the display region. Remove or label fixations shorter than a minimum predefined threshold and longer than a maximum predefined threshold.

3. Visually inspect data. Visually inspect data by using available visualization tools to ensure data validity. Check for signal loss, excessive noise, missing samples, or misaligned fixations. After each recording session, visually inspect the data using the visualization tools provided with the eye-tracking software (e.g., gaze plots, scanpaths, time-series traces). This step helps ensure that the data are of sufficient quality before proceeding to preprocessing and analysis.

1. In particular, check for signal loss and tracking problems by looking for long gaps in gaze patterns or regions on the stimuli where no gaze samples are recorded. Frequent or prolonged loss of tracking may indicate that the session should be excluded or repeated.
2. To check excessive noise, examine the raw x/y position traces for sudden jumps or erroneous samples that do not correspond to the stimulus layout. Excessive noise may indicate poor calibration, unstable head position, or hardware issues.
3. To spot missing or irregular samples, check whether the sampling rate is approximately as expected (e.g., 500 Hz) and whether there are unexpected clusters of missing samples beyond normal blinks.
4. To spot misaligned fixations, overlay fixations on the presented stimuli and verify that they fall on the intended lines and words. Systematic offsets (e.g., fixations consistently outside the text) may indicate calibration drift or an incorrect mapping between gaze coordinates and screen coordinates. The recorded gaze patterns should roughly follow the structure of the text (e.g., left-to-right progressions, line by line), with fixations clustering on words and relatively smooth saccades between them. A systematic deviation from this pattern, or frequent loss of usable data may require exclusion of the affected data from further analysis.

4. Perform drift correction. Perform vertical drift error correction where needed. Apply line-wise vertical drift correction for multiple-row text. Check the applicability of validated methods²¹.

NOTE: [Figure 1](#) presents a depiction of the steps involved in designing an eye-tracking experiment setup and data analysis pipeline, as described in the protocol section. The following section presents representative outcomes from recorded sessions.

Results

Data quality and validation

Successful sessions typically yield validation accuracy with mean error ≤ 0.5 – 1.0° and max/per-point error < 1.0 – 1.5° across targets; report both mean and per-point values. Provide a histogram of per-point error and a table summarizing mean $\pm$ SD and maximum error per session. These criteria align with common practice and vendor recommendations; sessions exceeding these thresholds should be partially or totally excluded, depending on study rules. [Figure 2](#) demonstrates a sample per-point error histogram for calibration accuracy for a 9-point calibration, involving three calibration attempts.

Precision and data loss

Report precision (e.g., RMS –Root Mean Square– sample-to-sample dispersion during fixation) and data loss (% missing samples per trial/session). Good trial runs usually show low dispersion at fixation and low data loss (e.g., < 5 – 10% , depending on the eye-tracking equipment and setup). Provide distributions of data-loss percentages and a by-trial scatter of precision vs. loss to reveal outliers. Check the current literature on definitions and minimal reporting items^{3,4,5,6,7,8,9,10,11}.

Drift behavior (pre/post correction)

Good trial runs exhibit small, near-constant vertical offsets that are substantially reduced by line-wise correction, either by automated drift-correction algorithms²¹ or manually. On the other hand, problematic runs show non-monotonic or large residual drift. Include median absolute vertical error pre/post correction. Check the applicability of automated drift-correction methods.

Distributions of common reading measures

Provide summary distributions for first fixation duration (FFD), single fixation duration (SFD), first-pass duration (FP), and total/gaze duration (GD), go-past time, skipping probability, and regression-in probability. For skilled adult reading of alphabetic text, fixation durations typically center around ~ 200 – 250 ms (with substantial skew), and saccades span ~ 7 – 9 characters on average. Values far outside these ranges or highly bimodal saccade distributions should be checked if they are indicative of expected progressive and regressive saccades due to task constraints or processing difficulties rather than timing errors. The choice and range of measures may vary accordingly across populations, specific languages, and orthographies. [Figure 3](#) shows a sample fixation duration distribution protocol from a representative session.

Sanity checks

An increase in word length (long words) usually leads to reduced skipping rate and longer gaze durations on those words. On the

other hand, function words (e.g., articles or prepositions) are usually skipped. Highly frequent words and highly predictable words based on the linguistic context (within the text or sentence) typically result in shorter first fixation and gaze durations, and higher skipping rates (the opposite is expected for low-frequency

and low-predictability words). Check if the data replicate some of these patterns, at least in one direction. Large reversals may indicate drift errors, misalignment of words and fixations, or data-processing errors.

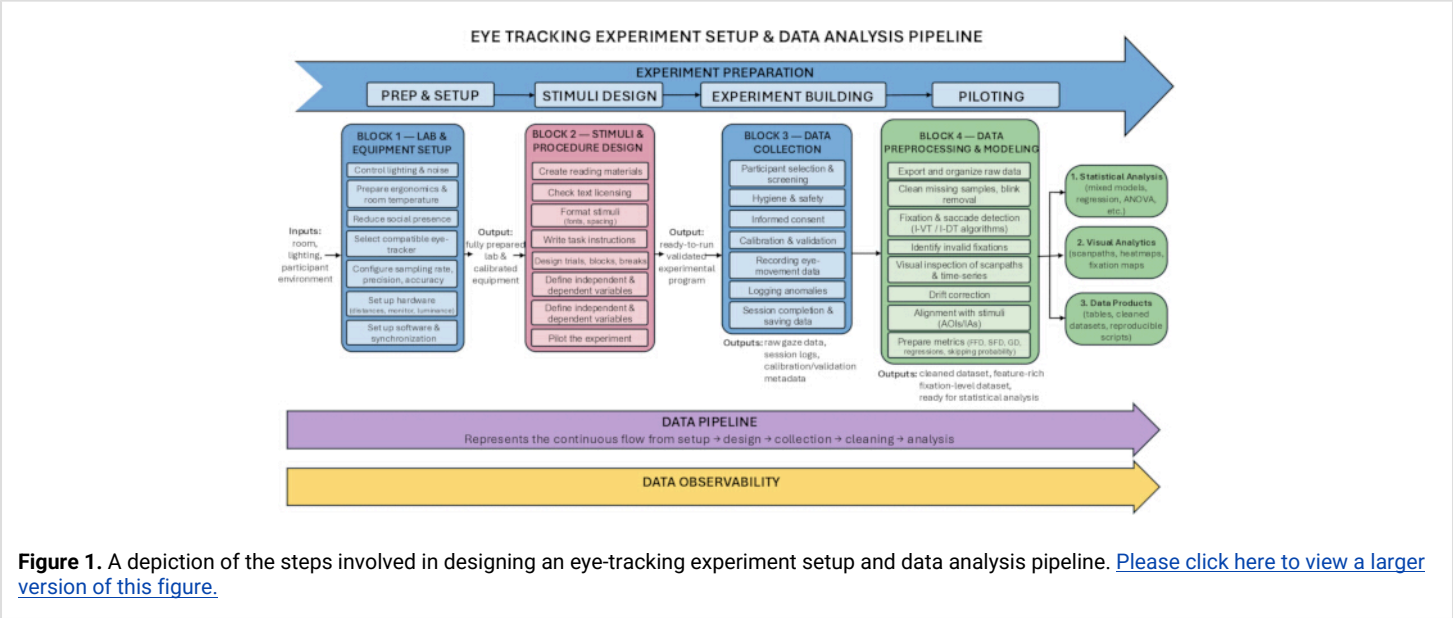


Figure 1. A depiction of the steps involved in designing an eye-tracking experiment setup and data analysis pipeline. [Please click here to view a larger version of this figure.](#)

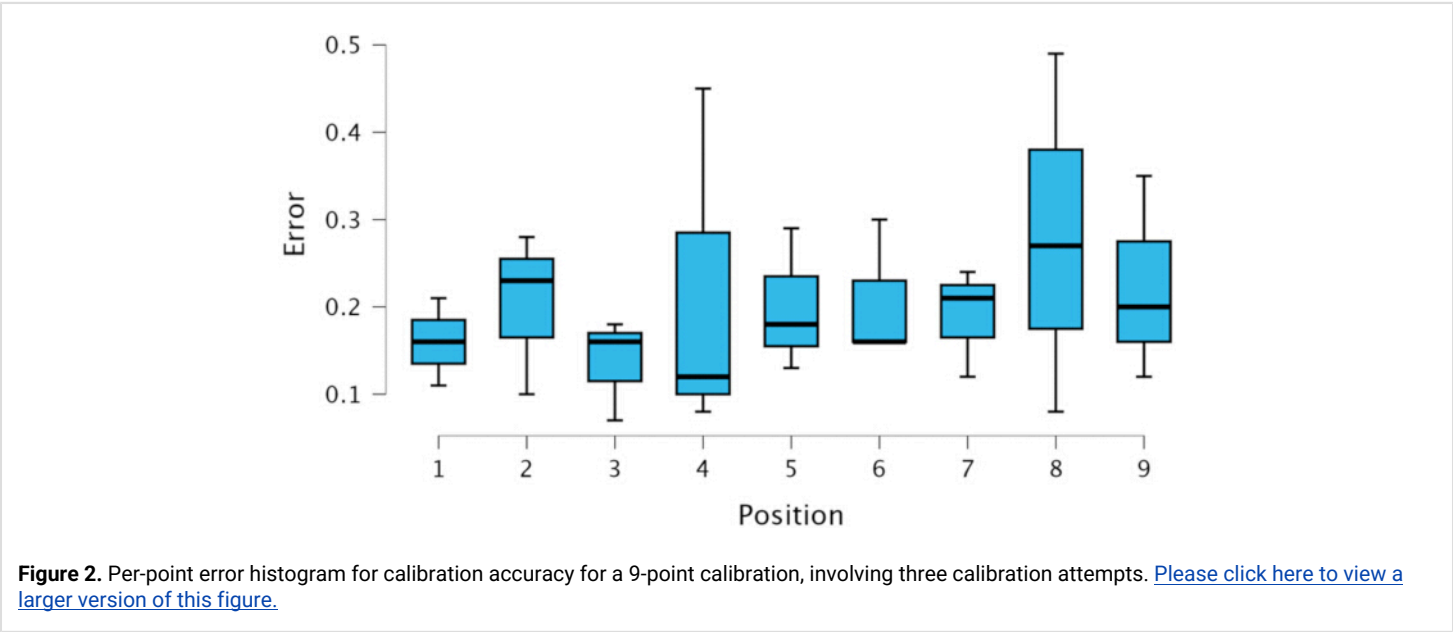


Figure 2. Per-point error histogram for calibration accuracy for a 9-point calibration, involving three calibration attempts. [Please click here to view a larger version of this figure.](#)



Figure 3. Fixation duration distribution protocol from a representative session; one participant reading an excerpt from the Little Prince novel. The position and the size of the circles show fixation locations and durations, respectively. The numbers next to the circles show fixation durations in ms. [Please click here to view a larger version of this figure.](#)

Discussion

The protocol presented here summarizes the main steps required to obtain reliable eye-tracking data in reading research. Although many of these steps may be well known to experienced researchers, they are often scattered across separate methodological papers, vendor documentation, and informal lab practices. By organizing them into a single workflow from lab preparation to preprocessing, we aim to provide a practical guide for researchers who are new to eye-tracking studies of reading. Compared to general-purpose eye-tracking guidelines or resources mostly focusing on eye-movement detection algorithms, this protocol provides an end-to-end framework as a response to reading research demanding data recording with high spatial resolution and sampling frequency.

A central theme across stages of the workflow is the importance of proactive data-quality management, starting with environmental control and equipment setup. Small issues such as reflections on the monitor, chair instability, or inconsistent viewing distances can produce systematic biases in fixation location or precision. Addressing these factors early minimizes the need for heavy post-hoc correction and supports more consistent validation accuracy (mean error ≤ 0.5 – 1°).

The protocol also emphasizes the importance of the alignment between research questions and hardware characteristics. Many studies in reading rely on precise spatial and temporal measurement of fixations, saccade landing positions, and regressions^{1,2}. Choosing an eye tracker with an appropriate sampling rate, precision, and accuracy is therefore essential, especially for measures such as landing position or parafoveal preview that are sensitive to small spatial errors. Depending on the specific study constraints, these methods can be modified.

For example, when working with neurodivergent populations who cannot tolerate a chinrest, researchers might shift to remote, head-free trackers and compensate for the loss of spatial precision by utilizing larger fonts, wider line spacing, and broader Interest Areas. When head-free systems are used, researchers must account for increased movement and potential data loss. Although remote trackers support more naturalistic setups, head stabilization still remains preferable for most reading tasks today.

Experiment design is another key factor that affects both data quality and interpretability. Standardized on-screen instructions, trial randomization, and appropriate block structure help maintain stable participant behavior across the session. Including drift checks at regular intervals is especially important in reading experiments, as even small gradual offsets can accumulate across lines of text. The protocol also supports the use of language-dependent lexical norms and transparent reporting of independent and dependent variables, which facilitates replication and comparison across studies.

During data collection, continuous monitoring of accuracy, posture, and participant comfort helps prevent large-scale data loss. Active troubleshooting is often necessary for common problems. For instance, calibration may fail at screen corners, then adjusting the monitor or chair height can resolve peripheral tracking loss. Issues from reflective glasses or contact lenses can sometimes be resolved by adjusting ambient lighting or changing monitor angles. Maintaining a detailed session log supports later troubleshooting and strengthens transparency by documenting interruptions, recalibrations, and unexpected events. This is increasingly important because many published studies now require explicit reporting of quality metrics, including precision and percentage data loss.

The preprocessing section emphasizes consistent and well-documented choices in event detection, drift correction, and

cleaning routines. Because fixation classification algorithms and thresholds can strongly influence the resulting measures, researchers should specify all parameter settings and justify them when they differ from common defaults. Visual inspection remains a critical step, even when automated tools are applied. The protocol also encourages reporting standard quality indicators such as validation accuracy, precision, data loss, and drift behavior, so that studies can be compared more reliably.

Moreover, the protocol recommends simple sanity checks of reading behavior, such as effects of word length, frequency, or predictability. Reproducing these canonical patterns provides an additional safeguard against unnoticed alignment issues, problematic drift, or preprocessing errors. These checks do not replace full statistical analysis but offer quick feedback on whether the data behaves as expected for skilled reading.

In addition to the details of the experimental protocol already mentioned, it is important to describe the linguistic and psychometric characteristics of the participants. The main rationale for this is the challenge experimental researchers face regarding the comparability of different studies. This is particularly important in the age of open science, where publicly available datasets may be used by researchers who did not originally collect the data and may therefore be unaware of potential idiosyncrasies arising from variations in participants' (socio)linguistic backgrounds.

Regarding linguistic background, the mother tongue of the participants should be noted, and information about the dialects they use should be collected to account for any dialectal differences that may impact the results. If participants speak more than one language – a common situation in today's world – second language proficiency should be recorded. In studies of bilingualism or second language processing, second language proficiency should be assessed using validated second language placement tests. Additionally, relevant aspects of socio-economic status may be recorded (e.g., participants' reading habits).

Tests should be conducted to assess psychometric characteristics of the participants, such as working memory span and cognitive flexibility (especially in bilingual studies). From a methodological perspective, one should consider which tests to select, their cognitive and time demands, and when to administer these measurements within the experimental protocol – whether before, after the eye-tracking session, or in a separate session.

It is essential to remember that in reading studies, language behavior is measured and explained using both the linguistic characteristics of the stimulus and the psychometric characteristics of the readers (or any other features relevant to the linguistic behavior of the participants)^{11,12,17,20}. To enable comparison across studies, all factors that may be relevant for data interpretation must be included. The choice of which

characteristics to include will depend on the purpose of the study. The data must be comparable. In summary, data should be collected and described in sufficient detail to allow any potential user the greatest possible flexibility in its use. These considerations may not be part of the protocol itself, but they are relevant for every individual study. In today's global scientific environment, such details increase the future potential use of the collected data.

This protocol has limitations, most notably the trade-off between data quality and ecological validity, as reading on a static monitor with a restrained head does not fully capture naturalistic reading behavior, and strict visual inclusion criteria can limit sample representativeness. Moreover, eye trackers differ substantially in coordinate scaling, timestamp precision, and event classification. A limitation of the current protocol is that it does not clarify how the variability affects reproducibility across laboratories or hardware brands. Creating guidance for cross-device validation, and for realistic future applications, such as adapting laboratory paradigms into immersive Virtual Reality or Augmented Reality reading environments, and developing digital biomarkers for cognitive or attentional disorders, would fill an important gap.

Overall, the protocol consolidates existing methodological guidance into a clear and transparent workflow for collecting high-quality eye-tracking data in reading. By following the steps and reporting the associated quality metrics, researchers can improve the reproducibility and credibility of reading studies and help establish stronger methodological consistency across the field.

Disclosures

NA

Acknowledgments

Funding: This project has been supported by COST–European Cooperation in Science and Technology (CA21131–Enabling multilingual eye-tracking data collection for human and machine language processing research (MultiPEYE), <https://www.cost.eu/actions/CA21131>)²⁴, The Foundation for Science and Technology, Portugal (grants: UID/214/2025, UI/BD/154500/2022), and Jagiellonian University Strategic Programme Excellence Initiative Priority Research Area DigiWorld (ID.UJ) project “Cognitive Aspects of Interaction and Communication in Natural and Artificial Agents”. We thank Anna Maria Wilkosz for their kind technical support.

References

1. Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychol Bull.* **124**, 372-422 (1998).

2. Rayner, K. Eye movements and attention in reading, scene perception, and visual search. *Q J Exp Psychol.* **62**, 1457-1506 (2009).
3. Carter, B. T., Luke, S. G. Best practices in eye tracking research. *Behav Res Methods.* **52**, 278-295 (2020).
4. Hessels, R. S. et al. The fundamentals of eye tracking part 1: The link between theory and research question. *Behav Res Methods.* **56**, 16 (2024).
5. Hooge, I. T. C. et al. The fundamentals of eye tracking part 2: From research question to operationalization. *Behav Res Methods.* **57**, 73 (2025).
6. Nyström, M. et al. The fundamentals of eye tracking part 3: How to choose an eye tracker. *Behav Res Methods.* **57**, 67 (2025).
7. Niehorster, D. C. et al. The fundamentals of eye tracking part 4: Tools for conducting an eye tracking study. *Behav Res Methods.* **57**, 46 (2025).
8. Hessels, R. S. et al. The fundamentals of eye tracking part 5: The importance of piloting. *Behav Res Methods.* **57**, 216 (2025).
9. Dunn, M. J. et al. Minimal reporting guideline for research involving eye tracking (2023 edition). *Behav Res Methods.* **56**, 4351-4357 (2024).
10. Eskenazi, M. A. Best practices for cleaning eye movement data in reading research. *Behav Res Methods.* **56**, 2083-2093 (2024).
11. Schotter, E. R., Dillon, B. A beginner's guide to eye tracking for psycholinguistic studies of reading. *Behav Res Methods.* **57**, 68 (2025).
12. Conklin, K., Pellicer-Sánchez, A. Using eye tracking in applied linguistics and second language research. *Second Lang. Res.* **32** (3), 453-468 (2016).
13. Pellicer-Sánchez, A. Eye-tracking in vocabulary research: Introduction to the special issue. *Res Methods Appl Linguist.* **3** (1), 100095 (2024).
14. Hutton, S. B. Eye-tracking methodology. In *Eye movement research: An introduction to its scientific foundations and applications*. Springer International Publishing, 277-308 (2019).
15. Pokhoday, M. Y. et al. Eye tracking methods in psycholinguistics and parallel EEG recording. *Neurosci Behav Physi.* **53**, 220-229 (2023).
16. Azman, H., Mihat, W., Soh, O. K. Analyzing stimuli presentations and exit-interview protocols to improve wearable eye-tracking data collection guidelines for reading research. *IJCALLT.* **11** (2), 51-65 (2021).
17. Papadopoulos, T. C. et al. Eye-tracking in reading research: A systematic review of studies with children of varying reading ability. *Eur Psychol.* doi: 10.1027/1016-9040/a000579 (2026).
18. Garlichs, A., Lustig, M., Gamer, M., Blank, H. Protocol to study how expectations guide predictive eye movements and information sampling in humans. *STAR Protoc.* **6** (2), 103737 (2025).
19. Moreno, J., León, J., Kaakinen, J. K., Hyönä. Relevance instructions combined with elaborative interrogation facilitate strategic reading: Evidence from eye movements. *J Psicol Educ.* **27** (1), 51-65 (2020).
20. Acartürk, C. *Eyes on Text: Eye movements in reading and language processing*. John Benjamins (2025).
21. Steinhauer, S. R., Bradley, M. M., Siegle, G. J., Roecklein, K. A., Dix, A. Publication guidelines and recommendations for pupillary measurement in psychophysiological studies. *Psychophysiology.* **59** (4), e14035 (2022).
22. Papesch, M. H., Goldinger, S. D. *Modern pupillometry: Cognition, neuroscience, and practical applications*. Springer (2024).
23. Salvucci, D. D., Goldberg, J. H. Identifying fixations and saccades in eye-tracking protocols. *Proc Symposium Eye Tracking Res Appl.* 71-78 (2000).
24. Jakobi, D. N. et al. MultiPLEYE: creating a multilingual eye-tracking-while-reading corpus. *Proc Symposium Eye Tracking Res Appl.* art. no. 111, ACM (2025).