Evaluating the data quality of the Gazepoint GP3 low-cost eye tracker when used independently by study participants

John Brand¹ · Solomon G. Diamond² · Natalie Thomas^{1,2} · Diane Gilbert-Diamond¹

Accepted: 17 October 2020 / Published online: 27 November 2020
© The Psychonomic Society, Inc. 2020

Abstract

The portability of low-cost eye trackers makes them attractive for research outside of the laboratory. Such research may require independent eye-tracker use. The present work compared the data quality of the Gazepoint GP3 when used independently by research participants with expert eye-tracking users. Twenty participants completed a training and a testing session 1 week apart. At training visits, participants were taught how to set up and use eye-tracking hardware and software and how to complete two tasks: a calibration task to measure accuracy and precision, as well as a visual search task to assess target fixations. At the testing session, participants set up the Gazepoint eye tracker and completed the two tasks without assistance. Participant accuracy and precision and visual search performance were compared to values obtained from two expert eye-tracking users. Additionally, the eye-tracker sampling rate, which is sensitive to factors such as head motion, was assessed in both participants and the expert users. Participant accuracy and precision closely approximated expert user values. Participant target fixations were detected with a 92.5% sensitivity and 76.8% specificity, closely mirroring expert user sensitivity and specificity. The sampling rate distribution was also similar between the participants and expert user (the means of those distributions were 16.99 ± 3.0 ms and 16.43 ± 2.3 ms, respectively). When used independently, data quality obtained from a low-cost, portable eye-tracking setup closely approximated values obtained from an expert user and was adequate enough to be a feasible option for some studies that require independent use by study participants.

Keywords Note: This data is mandatory

Introduction

Eye movement measurements are commonly used in psychological research to study cognitive processes, such as reading, scene perception, memory, and attention (Duc, Bays, Husain, Kennard, & Leigh, 2008; Kruijine & Meeter, 2016; Rayner, 2009). Traditional eye-tracking research has been conducted in laboratory settings, limiting its ecological validity. There is

increasing evidence of the importance of examining cognitive processes within their natural context (such as eye movements) in the real world (Ladouce, Donaldson, Dudchenko, & Ietswaart, 2017). Traditional eye-tracking setups do not allow for data collection outside of the laboratory because they are limited by high cost and immobility. Portable, low-cost eye trackers are now available that may allow traditionally lab-bound cognitive measurements to be assessed in non-lab environments (e.g., homes, schools, and the workplace) over extended periods of time.

The ability to collect data in non-laboratory settings may allow for new types of scientific investigations and offer a substantive shift from research that relies on nonrepresentative samples and small sample sizes. For example, it may be possible to reach a larger and more diverse population than the participants who are willing and able to travel to research settings (Konkel, 2015). Additionally, research conducted outside the laboratory may be less affected by demand characteristics because participants may act more naturally when in their normal environments (Ladouce et al., 2017). Remote eye tracking may be particularly beneficial to intervention

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13428-020-01504-2>.

✉ John Brand
John.brand@dartmouth.edu

¹ Department of Epidemiology, Geisel School of Medicine at Dartmouth College, HB 7927, One Medical Center Drive, Lebanon, NH 03765, USA

² Thayer School of Engineering at Dartmouth College, Hanover, NH, USA

research. Because of its high temporal and spatial resolution, ocular measures can assess cognitive processes in real time and are used to evaluate how people respond to task demands on a moment-by-moment basis. Because of this, eye-tracking paradigms have become a popular method to not only measure cognitive processing (Rayner, 2009), but also to deliver and evaluate interventions that target cognitive processes (Chen, Clarke, Watson, MacLeod, & Guastella, 2014; Field, Munafo, & Franken, 2009). However, because of the high participant burden associated with traveling to the lab for multiple training sessions, many of these studies are limited to single sessions, reducing their long-term effectiveness (Ferrari, Möbius, van Opdorp, Becker, & Rinck, 2016). If employed outside the laboratory, eye trackers may allow for better designed long-term interventions by reducing the need for multiple laboratory visits.

Data quality of commercial-grade eye trackers is understandably superior to that of low-cost systems (Funke et al., 2016). However, several studies have evaluated whether the data quality of low-cost eye trackers is suitable for use in a wide range of disciplines, with promising results. These assessments have shown low-cost systems are suitable for research relating to cognitive psychology (Čegovnik, Stojmenova, Jakus, & Sodnik, 2018), cartography (Wei, Popelka, Stachoň, Šašinka, & Doležalová, 2016), human computer interaction (Mantiuk, Kowalik, Nowosielski, & Bazyluk, 2012), information systems (Zugal & Pinggera, 2014), and the assessment and rehabilitation of motion disabilities (Gavas et al., 2018). However, all these assessments have been conducted in controlled laboratory studies where independent use was not required. The data quality of low-cost eye trackers when used independently by study participants has not received adequate scientific review for researchers to confidently apply the methodology in research projects. The present work aims to address this gap.

Measuring eye-tracking data quality

Eye trackers record eye movements by monitoring point of gaze, the current location on a screen at which an observer is looking at any given moment. The sample rate of an eye tracker determines the rate at which point-of-gaze measurements are recorded. For example, an eye tracker that records at 60 Hz records a gaze point approximately every 16.67 ms. Gaze point measurements that occur near each other in time and space are grouped to create a fixation, a period lasting approximately 20–300 ms in which the eyes are stationary. A fixation ends with a saccade, a rapid eye movement that moves the eyes to a new fixation. Data quality may be assessed by measuring the accuracy and precision of gaze point recordings. Accuracy is the ability of an eye tracker to measure where on the screen a participant is looking.

Precision is the ability of the eye tracker to produce consistent measurements ((Holmqvist et al., 2010; Holmqvist, Nyström, & Mulvey, 2012). Although it is common to evaluate accuracy and precision in isolation, it is also important to take into account their interaction. Figure 1 shows the possible combinations between precision and accuracy when fixating a target. A participant may be (1) accurate and precise, (2) accurate but not precise, (3) not accurate but precise, or (4) not accurate and not precise. The interaction between accuracy and precision may vary between participants and screen location, resulting in different data quality across participants and screen regions. Accuracy and precision are measured in degrees of visual angle. The lower the degree of visual angle, the better the accuracy or precision.

Data quality may be further assessed by evaluating sampling rate variability (Holmqvist et al., 2010). Manufacturers report a nominal sampling rate for their eye trackers that includes lower and upper limits based on a reported standard deviation. For example, the sampling range for a 60 ± 0.002 Hz eye tracker may range from 60.15 to 59.7 Hz (Tobii, 2010). In this case, it can be expected that a gaze point sample will be recorded every 16.63 to 16.75 ms. There are two main factors that may affect sampling rate variability: (1) the timing precision of the eye tracker and (2) the number of invalid samples.

Timing precision refers to the consistency of the eye tracker to record gaze points according to its specifications under ideal eye-tracking conditions. Timing precision is mostly affected by a computer's processing power. Complicated experiments require more processing power and may increase sampling rate variability. To ensure timing precision, manufacturers recommend the use of computers with specific specifications, such as those with a minimum processing speed and memory capacity. Controlling an eye tracker with a computer that does not meet specifications may cause delays in gaze point recordings that can cause increased timing variability that may lead to less robust eye-tracking measures (e.g., less precise measures of fixation durations). An invalid sample results when during a gaze point sample, a critical component of the eye needed for tracking cannot be reliably detected. Reasons for invalid samples may include, but are not limited to, a change in corneal reflection due to participant movement or a lighting change. Eye trackers flag invalid samples using validity codes in the data file. Because invalid samples are often removed prior to data analysis, sampling rate variability may be affected by a large number of invalid samples. For example, when using a 60 Hz eye tracker, the removal of each invalid sample would result in an average increase of 16.67 ms between samples that bookend the removed sample. Sampling rate variability may be assessed by computing the inter-sample time between valid samples measured in milliseconds, and then by visually inspecting the frequency distribution of inter-sample times. The frequency distribution

Fig. 1 Possible interactions between accuracy and precision. Dots represent individual point-of-gaze recordings that make up a fixation to the target cross. A fixation may be (a) accurate and precise, (b) not accurate but precise, (c) accurate but not precise, or (d) not accurate and not precise

should approximate a distribution based on the mean and standard deviation provided by the manufacturer.

It is also important to evaluate the ability of eye-tracking software to identify the location of fixations on a screen because of their importance as a dependent variable in psychological research (Dalmajer, 2014). Fixation identification may be evaluated using a task that requires participants to make a series of fixations where a ground truth for the fixation can be identified. One such task is the visual search task. Visual search tasks require participants to make a series of saccades and fixations to locate a target object among a set of distractors. By asking participants to indicate when they are fixating a target using a button press, visual search performance may be used to evaluate the ability of an eye tracker to identify fixations to target objects.

Assessing the data quality of low-cost eye-tracking systems

An eye tracker's data quality is typically assessed by its manufacturer using their propriety software. However, this quality assessment is often not applicable to experimental research because manufacturers rely on the use of motionless artificial eyes in ideal eye tracking conditions (Wang, Mulvey, Pelz, & Holmqvist, 2017). Data quality can be evaluated in comparison to that obtained by an expert eye-tracker user. Because of their expertise, an expert user should be able to closely

approximate the optimal data quality of an eye tracker under certain experimental conditions. Obtaining optimal eye-tracking data is affected more by correct eye-tracking setup and maintaining the correct setup throughout the duration of the study than by device characteristics (e.g., its sampling rate) or the controlling software (Ooms, Dupont, Lapon, & Popelka, 2015), the most important of which is maintaining correct distance from the eye tracker following correct device placement. Manufacturers report an ideal distance that the eye tracker should be placed from the user to obtain good data quality, usually ranging between 50 and 75 cm. Fortunately, most software also records associated eye position data with every gaze point observation, such as distance from the screen that can be used to evaluate how well participants maintain proper distance throughout the experiment. Any data quality evaluation should also evaluate whether participants can maintain proper distance throughout data collection.

The use of low-cost eye trackers in research settings often requires integration with customized software to control the eye tracker and accurately link it with stimulus presentations. To facilitate this integration, researchers have developed open-source software. Examples include PsychoPy (Peirce, 2007) and Pygaze (Dalmajer, Mathôt, & Van der Stigchel, 2014). It is necessary to evaluate data quality when using open-source software because each use customized algorithms to control eye-tracker events that are used to assess data quality. These events include logging information (e.g., stimulus onset and offset), real-time sampling of gaze position, and online detection of saccades and fixations. Additionally,

when compared to a manufacturer's reported sampling rate, variability may increase when using third-party software because of the increased computational demands associated with creating complex cognitive experiments (e.g., studies with precise stimuli generation and timing needs).

Among the open-source options, Pygaze is best suited for eye-tracking experiments in naturalistic settings. PyGaze-based plugins are available for the freeware OpenSesame (Mathôt, Schreij, & Theeuwes, 2012), an easy-to-use graphical experiment builder that is capable of customizing text, visual, audio, and video stimuli. Additionally, PyGaze has been used to assess data quality of a low-cost system, the Eye Tribe (Dalmaijer, 2014). That study showed precision and accuracy below 0.5° of visual angle, as well as inter-sample times around the expected sampling rate and accurate fixation detection. However, the study was conducted under optimal eye-tracking conditions including use of a chin rest to limit head movement and experienced eye-tracking participants, so the results may not apply to a less controlled environment with naïve eye-tracking participants. Thus, prior to using Pygaze software with a low-cost eye tracker in nontraditional research settings, a suitable device that is simple to set up should be selected and tested for data quality.

Currently, one of the few manufacturers of low-cost eye trackers is Gazepoint (Vancouver, British Columbia, Canada). Gazepoint eye trackers offer a large number of eye-tracking parameters for a low-cost system. Among the most important parameters for researchers include point-of-gaze data, pupil dilation, cursor position, calibration data, head position measures, and a validity measure for each variable. Their most affordable model, the GP3, has a reported sampling rate of 16.67 ± 1.42 ms (Gazepoint, personal communication, July 22, 2020) and an accuracy rate between 0.5° and 1° of visual angle. However, they do not report estimated precision values. A setup that includes a Gazepoint eye tracker controlled by OpenSesame and Pygaze software may provide the ease of use necessary to collect robust data when used independently.

There are two major issues that may limit the independent use of low-cost eye trackers in non-laboratory environments: (1) the ease of eye-tracking setup (e.g., position of the eye tracker relative to the screen) and (2) the user-friendliness of the software (e.g., calibration procedure). Both issues can lead to misleading interpretations and outcomes. Gazepoint eye trackers require a nontechnical setup and offer both desktop and laptop configurations. Critically, Gazepoint's control software includes a live video feed that aids participants in the proper setup procedure. This is particularly beneficial given that correct device placement is unique to each individual, and the part of the setup procedure where participants are most likely to fail. Pygaze software compliments Gazepoint's simple setup by simplifying experiment execution by controlling all runtime operations, including calibration and data

collection, using a single click via an OpenSesame shortcut. This avoids the need to open OpenSesame's graphical user interface where it is possible to mistakenly alter the experiment code. Importantly, Pygaze provides immediate visual feedback of calibration results and may be programmed to automatically repeat calibration if results do not meet a certain threshold. Because calibration is largely dependent on correct device placement, and data quality is largely dependent on good calibration results (Nyström, Andersson, Holmqvist, & van de Weijer, 2013), this built-in feature provides a further check to ensure robust data collection.

Additionally, the Gazepoint GP3 has been utilized in a wide range of investigations that may benefit from independent participant use. Applications include those with educational relevance, including real-time assessments of engagement and performance during computerized courses (Kim, Pollanen, Reynolds, & Burr, 2020; Notaro & Diamond, 2018), as well as tracking student performance over time (Bottos & Balasingam, 2019; Calado, Marcelino-Jesus, Luis-Ferreira, & Sarraipa, 2019). Other related uses have been described for improving the measurement of a variety of cognitive processes, including attention (Brand, Masterson, Emond, Lansigan, & Gilbert-Diamond, 2020; Taylor, Bilgrien, He, & Siegelmann, 2015) and memory (Taylor et al., 2015). Of particular relevance is a series of studies that have measured cognitive load during a series of complicated tasks using pupillometry (Sibley et al., 2021; Sibley, Foroughi, Brown, & Coyne, 2018; Sibley, Foroughi, Olson, Moclaire, & Coyne, 2017). These studies were conducted in a group classroom setting where participants were required to independently complete the gaze calibration process but were not required to complete any hardware setup. Despite this promising research, there has been little overall device assessment, identification of device challenges, or guidance for GP3 device usage, specifically when used independently.

In the present work, we evaluated the data quality of the Gazepoint GP3 eye tracker when controlled using Pygaze and used independently by study participants. We asked participants to come to the laboratory on two separate occasions to complete two tasks: a calibration task to measure accuracy and precision, and a visual search task so we could assess target fixations. The first visit was a training visit. Individuals were shown how to use the eye-tracking hardware and software. At the second visit, participants were provided a laptop and Gazepoint device and asked to set up and complete the calibration and visual search tasks as taught in visit 1. Accuracy and precision, sampling rate variability, and visual search performance of participants at time 2 was compared to measures obtained by two expert eye-tracker users. Additionally, we evaluated whether participants were able to maintain correct distance with the eye tracker in real time. Given the simplicity of the Gazepoint setup, we predicted that the data quality

obtained from study participants using the device independently would approximate those of an expert user.

Methods

Participants Twenty-two undergraduate and graduate students completed the study in exchange for monetary compensation. Of the 22, two completed the training session but not the testing session. Thus, a total of 20 were included in the final analyses. Of those, 15 were female (75.0%), and the mean age was 21.86 years ($SD = 5.35$). Three people reported wearing contacts and two people reported wearing glasses. The remaining 15 reported uncorrected normal vision. All provided written informed consent, and all study procedures were approved by the Committee for the Protection for Human Subjects at Dartmouth College.

Tasks

Calibration task. Participants were presented with 12 black dots appearing consecutively, equally spaced on a 4×3 target grid. The target grid spanned 34.5° of visual angle horizontally and 19.41° vertically. Each fixation dot had a diameter of 0.44° . Dots were displayed individually for 2000 ms, followed by an inter-target interval of 250 ms in which the screen was blank. Participants were instructed to fixate on the targets until a new target appeared.

Visual search task. The visual search task was a replication of previous work (Kamienkowski, Ison, Quiroga, & Sigman, 2012). On each trial, participants were presented with two target stimuli and 16 distractor stimuli, presented at random locations. Targets were the letter “E” bordered by eight “#” symbols. Distractors were a mirrored “E” bordered by eight “#” symbols. The size of each target and each distractor was

55×75 pixels, excluding the surrounding symbol border. When including the surrounding border, the size of each target and distractor was 165×245 pixels. The minimum spacing between neighboring stimuli was 150 pixels, and no stimuli appeared within 150 pixels of the bottom of the screen. This increased the probability that all stimuli were visible to participants despite the eye-tracking mount occluding the bottom-most part of the screen. When viewed from a distance of 57 cm, the size of the stimuli subtended a visual angle of 0.9° without the border and 4.5° with the border. All stimuli were gray-scaled and presented on a white background. These stimuli were selected because the # symbol border minimizes the probability that people can detect targets using peripheral vision.

Participants were instructed to locate the two targets as quickly as possible. To indicate they found a target, participants were instructed to click the mouse while fixating the target center. A blank screen was shown for 250 ms following localization of the second target. A new trial started with the presentation of a new search array. An example search array is shown in Fig. 2.

Eye-tracker and computer setup Eye movements were recorded using the Gazepoint GP3 eye tracker recording at 60 Hz. The eye tracker was connected to an HP laptop (Hewlett-Packard, Palo Alto, CA, USA) at a screen resolution of 1366×768 and refresh rate of 60 Hz using Gazepoint’s laptop mount. OpenSesame software (Mathôt et al., 2012) was used to display the stimuli and controlled all timing and response operations, including eye-tracker integration using PyGaze software (E. S. Dalmaijer, Mathôt, & Van der Stigchel, 2014).

Procedure Participants had two lab visits approximately 1 week apart. The first session was an individualized training session in which participants were taught proper eye-tracking setup and use by the first author. Upon arrival, participants were seated and shown the eye-tracker setup, including proper

Fig. 2 (a) Laptop setup. Screen locations were constrained so that no stimuli appeared within 150 pixels of the bottom of the screen, increasing the possibility that all stimuli were visible to the participant. (b) Example target and distractor stimuli

placement and computer connection (shown in Fig. 2). After setup, participants were instructed to open the Gazepoint control software, which displays a real-time video feed that can be used to verify eye capture and correct distance from the screen. Successful eye capture is indicated by green boxes bordering the left and right eye. Distance from the screen is assessed by a single dot that moves horizontally across the screen; turning red if the participant is too far and yellow if the participant is too close. A green dot in the middle signals correct distance. Participants were instructed to verify that their eyes were centered in the image and were being captured, and that the green dot was positioned in the center, indicating correct distance. If one or more of these criteria were not met, participants were instructed to adjust the camera or their body position. After verifying correct positioning, participants were asked to minimize head and body movement for the duration of the study.

Once positioned correctly, participants then started the experiment by clicking an OpenSesame shortcut, avoiding the need to open OpenSesame's graphic user interface. Participants then completed an abridged version of the experimental protocol that included the calibration task and 10 trials of the visual search task. Once completed, participants were instructed to disassemble the eye tracker, repeat the setup procedure and then run themselves through a full version of the experimental protocol. The full experimental protocol required participants to complete the calibration task and 50 trials of the visual search task. Once the full protocol was finished, participants were asked to disassemble the eye tracker. In both the abridged version and the experimental version, participants completed PyGaze's standard 12-point calibration prior to completing the tasks.

During visit 2, participants were given the same laptop, Gazepoint, and an instruction sheet, and were asked to set up the eye tracker and complete the experimental protocol as done in the training session. As in session 1, participants first completed the abridged version of the tasks to re-familiarize them to the study design and then completed the experimental protocol. However, unlike visit 1, participants were not required to disassemble the eye tracker between the abridged and full experimental protocol, although participants were instructed to ensure proper setup, including eye capture and distance from the screen. Participants during visit 2 were not given any instructions other than the written protocol. Only data collected at time 2 was analyzed. Data collected using the abridged protocol was not analyzed.

Preprocessing and data analysis OpenSesame outputs three data files: (1) an eye-tracking file that contains eye-tracking metrics outputted by Gazepoint's application programming interface (API), (2) a log file that contains calibration information calculated by PyGaze and messages sent to the eye tracker, and (3) a behavioral file that contains stimuli location

information. The following eye-tracking parameters were used in the present experiment: the best point of gaze in the horizontal (BPOGX) and vertical (BPOGY) directions, their associated validity metric (BPOGV), the timestamp vector (TIME), and the user-defined variable (USER). The USER variable was coded to include task and trial identification, stimuli location, and responses from peripheral devices (e.g., mouse clicks). The BPOGX and the BPOGY are the average of the left and right eye gaze points (in the x and y directions) if both eyes are available, or if not, then the value of either the left or right eye, whichever is available. The TIME vector records the time in microseconds since the beginning of the experiment, which is useful for synchronization with third-party software and to assess sampling rate variability. Distance from the screen was extracted using Gazepoint's LEYEZ and REYEZ variables, which records the distance between the eye-tracking camera and the left and right eyes on the z -axis.

Distance monitoring Gazepoint recommends that participants sit 65 ± 15 cm from the screen. Distance from the eye tracker throughout the experiment was examined for each participant to determine whether individuals remained within the manufacturer's suggested depth range throughout the study. To compute this metric, invalid LEYEZ and REYEZ observations were first removed using the BPGOV filter and then averaged together to obtain a central distance value. We also evaluated whether participant distance was associated with data quality, including accuracy and precision metrics, and the number of invalid samples. There were no significant associations between screen distance and any of these measures, so this factor was not included in any of our analyses.

Gaze point samples and sampling rate variability Prior to evaluating sampling rate variability, we first investigated the underlining sample data. Specifically, we counted the number of invalid samples using the BPOGX and BPOGY validity filters. We then counted the number of valid samples that fell within the boundaries of the screen during the calibration and visual search tasks.

The Gazepoint GP3 advertises a sampling rate variability of 16.67 ± 1.42 ms. Sampling rate variability within our sample was assessed by removing invalid samples, and then by computing the frequency distribution of valid samples in both the calibration and visual search tasks. For this analysis, inter-sample times were grouped into 0.0001 ms bins. The final bin was coded to include all sample times greater than 33.34 ms, equal to two consecutive sample observations.

Fixation accuracy and precision Horizontal accuracy, vertical accuracy, and horizontal and vertical precision were calculated for each participant at each target location during the calibration task. Accuracies were calculated by first subtracting

individual X (BPOGX) and Y (BPOGY) gaze points from the actual x and y target locations when a participant was fixating a target dot, excluding invalid samples. Mean horizontal and vertical accuracies were then computed for each participant and each target location. Horizontal and vertical precision for each participant at each target location was calculated by computing the horizontal and vertical root mean square (RMS) over all BPOGX and BPOGY samples obtained when a participant was fixating the target stimuli, excluding invalid observations. Prior to computing accuracy and precision metrics, the first 100 ms after dot onsets were pruned. This was done to allow participants time to fixate on new targets and to limit the effect of saccade landing error on accuracy and precision measurements.

Overall accuracy and precision, as well as in the x and y directions, were compared to expert values ($\alpha = 0.05$, two-tailed). We then compared accuracy and precision metrics across the different screen locations using separate multivariate analysis of variance (MANOVA) tests. For this analysis, we included participant data only, and accuracy and precision metrics were averaged over the x and y directions. Tukey's honestly significant difference (HSD) test was used to investigate pairwise comparisons between the individual target locations.

Visual search fixation identification The visual search task was used to assess the detection of target fixations. A target fixation was defined when the eye tracker detected a fixation within and area of interest (AOI) surrounding a target. PyGaze provides a simple algorithm for fixation detection but recommends that a more robust measure be used for optimal results. Therefore, we identified fixations using the saccades package in R. This package identifies fixations using a velocity-based algorithm developed by Engbert & Kliegl (2003). A total of eight AOIs were created. The first two were based on screen size: a 0×0 -pixel rectangle and a 1388×768 -pixel rectangle. The next two were based on target size: a 55×75 -pixel rectangle that corresponded to the size of the target, excluding the surrounding border, and a 165×245 -pixel rectangle that correspond to the size of the target, including the surrounding border. The final four were created in increments of 50 pixels, resulting in a 215×295 -pixel rectangle, a 265×345 -pixel rectangle, a 305×395 -pixel rectangle, and a 355×445 -pixel rectangle. Fixations were identified using minimum fixation durations of 100 and 150 ms, values suggested to be the theoretical minimum duration for a single fixation (Irwin, 1992; Manor & Gordon, 2003). We also included a 275 ms minimum fixation because this value represents the mean fixation duration for a visual search task (Rayner, 2009). The sensitivity and specificity to detect fixations on targets at each AOI for each fixation duration was then assessed by calculating a receiver operating characteristic (ROC) curve and computing the area under the curve (AUC).

The ROC curve was created by calculating true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) for each AOI for each participant on each trial. The ground truth for observing the target was defined by the mouse click. True positives were defined when a fixation was detected within a target AOI and that fixation was followed by a mouse press within 500 ms. False positives were defined when a fixation was detected within a target AOI, but that fixation was not followed by a mouse press within 500 ms. False negatives were defined when a fixation was not detected within a target AOI during the 500 ms proceeding a mouse click. True negatives were defined when a fixation was detected outside of the target AOI and that fixation was not followed by a mouse press within 500 ms. The numbers of TPs, FPs, FNs, and TNs were then summed across participants and used to calculate sensitivity and specificity measures. Sensitivity was calculated using the formula, $\frac{TP}{TP+FN}$. Specificity was calculated using the formula $\frac{TN}{TN+FP}$.

Expert data The first (JB) and third (NT) authors completed the testing protocol five times each to obtain comparison measures of accuracy and precision, sampling rate variability, and visual search performance. Each author completed individual testing sessions on a single day, with at least one hour between each session. The expert user measures were calculated with the same methods as the participant measures and averaged across individual testing sessions.

Results

Distance monitoring Participant distance from the eye tracker throughout the experiment is shown in Fig. 3. With the exception of a single participant, all participants remained within the manufacturer's reported depth range of 65 ± 15 cm throughout the experiment. The average mean and standard deviation across all participants were similar across the calibration (58.47 ± 5.77 cm) and visual search tasks ($M = 58.77 \pm 5.87$ cm). The average mean and standard deviation throughout the experiment were 58.7 ± 5.86 cm.

Gaze point samples and sampling rate variability There was a total of 398,199 gaze point observations recorded for participants, of which 386,235 (96.99%) were valid. The percent of valid samples was lower in the calibration ($27,431/29,828 = 91.96\%$) than visual search task ($358,804/368,371 = 97.40\%$). Of the valid samples, 97.3% were recorded within the screen boundaries during the calibration task and 96.7% during the visual search task. The total number of expert samples was 209,330, of which 206,776 were valid (98.79%). The percent of valid trials was similar between the calibration ($14,170/14,791 = 95.80\%$) and visual search ($192,606/194,539 =$

Fig. 3 Distance between the eye tracker and the recorded pupil. The black dotted line represents the recommended distance from the screen (65 cm). The upper dotted purple line represents the upper screen distance recommendation (75 cm), and the lower purple line represents the lower screen distance recommendation (50 cm). Red lines represent distance

from the screen during the calibration task, and green lines represent distance from the screen during the visual search task. All participants, with the exception of #14, remained within the manufacturer's reported depth range of 65 ± 15 cm throughout the experiment

99.00%) tasks. Approximately 95.7% of valid gaze points were recorded onscreen during the calibration task and 97.1% during the visual search task.

Expert and participant inter-sample distributions are shown Fig. 4. The average inter-sample time for participants was 16.99 ± 3.0 ms. Importantly, less than 0.05% of inter-sample times were greater than 33.34 ms. The average inter-sample time for the expert user was 16.43 ± 2.3 ms. There were no

sample times that exceeded 33.34 ms in the experienced eye-tracker data.

Fixation accuracy and precision Participant accuracy and precision metrics at each target location are visualized in Fig. 5. Visualizations for each participant are displayed in Supplementary Figure S1. Covariance ellipses represent the covariance matrix computed over all gaze points at each

Fig. 4 Density distributions of inter-sample times for (a) participants and (b) expert baseline data. Inter-sample times are grouped into 0.0001 s bins, with the final bin capturing all inter-sample times that exceed 34.34 ms (two consecutive samples)

Fig. 5 Participant point-of-gaze samples to each fixation dot location averaged across participants. Ellipses represent the covariance matrix to capture 95% of gaze points to a fixation dot. The horizontal and vertical

lengths of the ellipse represent the horizontal and vertical precision at each point. The black border represents the size of the screen, 1366×768 pixels

location to capture 95% of gaze points to a fixated target. The horizontal and vertical lengths of the ellipse represent the horizontal and vertical precision at each point. Participant and expert mean accuracy and average precision values the x and y directions are shown in Tables 1 and 2.

Overall, participant accuracy in the x - and y -direction were within Gazepoint's stated range of 0.5 – 1.0° of visual angle. When collapsed across the x - and y -direction, participant accuracy was statistically lower than expert accuracy of $0.81 \pm 0.17^\circ$ ($M = 0.70 \pm 0.12^\circ$; $P = 0.009$). Participant accuracy in the x -direction was similar to the expert value of $0.82 \pm 0.20^\circ$ ($M = 0.79 \pm 0.20^\circ$; $P = 0.572$), and accuracy in the y -direction was lower than the expert value of $0.80 \pm 0.23^\circ$ ($M = 0.62 \pm 0.12^\circ$; $P < 0.001$). Like accuracy, participant

precision collapsed across the x - and y -direction was within 0.5 – 1.0° of visual angle and was lower than the expert value of $1.01 \pm 0.64^\circ$ ($M = 0.81 \pm 0.17^\circ$; $P < 0.001$). Participant accuracy in the x direction was also lower than the expert value of $1.09 \pm 0.46^\circ$ ($M = 0.88 \pm 0.18^\circ$; $P = 0.002$), and precision in the y -direction was lower than the expert value of $0.93 \pm 0.88^\circ$ ($M = 0.74 \pm 0.18^\circ$; $P = 0.003$). Additionally, when collapsed across the x - and y -direction, accuracy and precision did not differ significantly between any of the screen locations.

Visual search fixation identification Sensitivity and specificity to target fixations were calculated to determine expert and participant ROC curves for fixation durations of 100, 150,

Table 1 Accuracy and precision in degrees of visual angle averaged across participants at each fixation dot location

Column	Horizontal accuracy				Mean	Vertical accuracy				Mean
	1	2	3	4		1	2	3	4	
Row										
1	1.05	0.49	0.52	1.00		0.65	0.77	0.69	0.71	
2	0.97	0.68	0.65	0.65		0.52	0.57	0.46	0.42	
3	1.03	0.69	0.81	0.90		0.62	0.71	0.63	0.72	
Average					0.79					0.62
Column	Horizontal precision				Mean	Vertical precision				Mean
	1	2	3	4		1	2	3	4	
Row										
1	1.08	0.68	0.60	1.00		0.70	0.78	0.66	0.67	
2	1.09	0.68	0.69	0.91		0.51	0.84	0.84	0.51	
3	1.12	0.86	0.80	1.03	0.88	0.64	0.71	1.13	0.90	0.74

Table 2 Baseline accuracy and precision in degrees of visual angle averaged across runs at each fixation dot location

	Horizontal accuracy					Vertical accuracy				
Column	1	2	3	4		1	2	3	4	
Row					Mean					Mean
1	0.92	0.70	0.86	1.05		0.67	0.90	1.13	1.29	
2	1.15	0.68	0.61	0.88		0.70	0.70	0.81	0.89	
3	0.87	0.42	0.79	0.94		0.46	0.56	0.79	0.65	
Average					0.82					0.80
	Horizontal precision					Vertical precision				
Column	1	2	3	4		1	2	3	4	
Row					Mean					Mean
1	1.24	0.67	2.28	1.42		0.64	0.80	3.64	0.85	
2	1.41	1.06	0.72	0.92		0.59	1.14	0.91	0.57	
3	0.92	0.68	0.86	0.95	1.09	0.31	0.57	0.66	0.46	0.93

and 275 ms. ROC curves for experts and averaged across participants for a fixation duration of 100 ms are shown in Fig. 6. Overall, there was good discriminability to detect targets using a fixation duration of 100 ms. The AOI corresponding to the size of the target including the surrounding border (AOI₃) produced the best tradeoff between specificity and sensitivity

for both participant and expert data. Participant sensitivity (92.5%) and specificity (76.8%) were similar to expert values of 85.5% and 82.8%. The results were unchanged when using fixation durations of 150 and 275 ms. The data for these durations as well as individual participant ROC curves can be found in the [Supplementary material](#).

AOI₁: 0 X 0-pixel rectangular AOI.

AOI₂: 55 X 75-pixel rectangular AOI (target size excluding surrounding border).

AOI₃: 165 X 245-pixel rectangular AOI (target size including surrounding border).

AOI₄: 215 X 295-pixel rectangular AOI.

AOI₅: 265 X 345-pixel rectangular AOI.

AOI₆: 305 X 395-pixel rectangular AOI.

AOI₇: 355 X 445-pixel rectangular AOI.

AOI₈: 1366 X 768-pixel rectangular AOI (screen size)

Fig. 6 ROC curves for target fixations with a minimum fixation duration of 100 ms averaged across (a) participants and (b) expert runs. AOIs surrounding the target image are color-coded to match the

corresponding sensitivity and specificity points on the graphs and the AOI description in the footnote

Discussion

The present study evaluated the data quality of the Gazepoint GP3 eye tracker controlled using PyGaze software and used independently by study participants. With the exception of a single participant, all participants remained an appropriate distance away from the eye tracker throughout the experiment. Importantly eye-tracker distance was not associated with any data quality measures, suggesting participants were successful in placing the device. Overall, participant accuracy and precision were within the manufacturer's stated guidelines of between 0.5° and 1° of visual angle and similar to or better than measures recorded by two expert users. Additionally, the number of valid samples was similar between participants (96.99%) and experts (98.79), and over 97% of gaze points were recorded within screen boundaries. Furthermore, the mean participant inter-sample time was 16.99 ± 3.0 ms, approximating the advertised sampling rate of 16.67 ± 1.4 ms and the expert value of 16.43 ± 2.3 ms. Additionally, a robust fixation algorithm was accurate in detecting target fixations during a visual search task. When using an AOI equal to target size (including its border), participant sensitivity to detect target fixations was 92.5% with a specificity rate of 76.8%. This closely matched expert sensitivity of 85.5%, with specificity of 83%. Together, these results suggest that when participants use the Gazepoint GP3 eye tracker independently, the data quality obtained using Pygaze software is comparable to the data quality obtained by expert eye-tracking observers.

This low-cost setup produced good data quality difficult eye-tracking conditions, including participant-driven setup, head free-to-move remote mode, and use of a laptop mount. This may not be surprising given the completed tasks assessed fixations that do not require a high sampling rate nor optimal eye-tracking conditions. Nevertheless, the present results are encouraging for researchers who wish to conduct non-laboratory experiments that require independent eye-tracking use by study participants. However, researchers should be aware that this suggestion applies only to studies that measure fixations. Saccades are too rapid to be accurately measured at 60 Hz (Dalmajer, 2014), and because of this we did not attempt to measure these eye movements. Future research is needed to examine whether low-cost eye trackers with increased sampling rates are able to accurately measure saccades. One candidate is the Gazepoint GP HD3 which offers an industry-leading 150 Hz for a low-cost system. We also did not measure pupillometry. Work with low-cost eye trackers—including the Gazepoint GP3—has shown that pupil measurements may be a reliable measure for some cognitive processes (e.g., cognitive workload) (Coyne & Sibley, 2016) but not others (e.g., working memory) (Sibley et al., 2018). However, given that environmental factors (e.g., lighting levels) are known to affect pupil dilation (Larson & Behrends, 2015), future research is needed to evaluate

whether the Gazepoint eye tracker accurately records pupil size outside of controlled laboratory settings.

It is encouraging that both participant (16.99 ± 3.0 ms) and expert data (16.43 ± 2.3 ms) approximated the advertised sampling rate of 16.67 ± 1.4 ms. However, it is important to note that both standard deviations were greater than the advertised value. This may be due to differing processing demands between controlling software. Gazepoint's camera has a strict fixed sampling rate, but the processing time for loading and displaying stimuli, transferring eye data from the hardware to software, and then logging the results depends on the processing demands of the experiment. The standard deviation reported by Gazepoint was recorded using their own software during a low processing demand task: a 15 s screen capture task using a preloaded image (Gazepoint, personal communication, July 22, 2020). Our task may have exhibited a larger SD, as the Pygaze software required increased processing power to control all timing and response operations, including online stimulus generation on a trial-to-trial basis over a longer period of time. Researchers should be aware that increased task complexity often requires more processing power to execute and may result in increased sampling rate variability.

Participants had a greater percent of invalid samples than the expert user. Once removed, this contributed to an increase in the average amount of time between valid samples and resulted in 0.5% of participant inter-sample times being equal to or larger than 32.34 ms (two consecutive samples). There were no expert-user inter-sample times that exceeded 32.34 ms. An invalid sample may be attributed to participant properties (e.g., fixation ability, pupil occlusion, retinal reflection quality), test environment properties (e.g., a sudden change in illumination levels), or a failure to maintain proper positioning (Tobii, 2011). However, it is often not possible to retroactively attribute a reason to an invalid sample (Holmqvist, Nyström, & Mulvey, 2012), making it difficult to explain differences between participants and the expert user. Nevertheless, even after removing invalid samples, this eye-tracking setup may have the timing precision to investigate a variety of cognitive processes. For example, target identification during visual search tasks requires measuring fixations between 180 and 275 ms, whereas scene perception and reading require measuring fixations in the 260–330 ms and 225–250 ms ranges, respectively (Rayner, 2009).

Because the laptop mount partially occluded the lower portion of the screen, we constrained our possible stimulus locations so that no stimuli appeared within 150 pixels of the bottom. This resulted in equal data quality across screen locations. Researchers, particularly those using a portable laptop setup, should be aware of possible viewing obstructions and plan studies accordingly. Furthermore, the placement of the laptop mount limits the usability of the keyboard, restricting possible participant responses. For this reason, eye-tracking studies using portable laptop setups may best suited for studies

where the need for a participant response is limited to a single key press of an unobstructed button (e.g., the spacebar), or where responses can be made using peripheral devices, such as a mouse, button box, joystick, or gamepad.

The present work has the following limitations. Because our task was short, we were not able to examine differences in data quality over time. It is common for data quality to degrade over the course of an eye-tracking experiment. Although not a direct study of data quality, a study using a Gazepoint eye tracker controlled by their propriety software showed that task complexity, rather than task duration, was associated with an increase in the number of invalid samples (Sibley et al., 2017). Depending on task complexity, it may be beneficial for researchers to build in accuracy and precision assessments during specific time points (such as natural breaks between tasks or blocks). This may benefit studies outside of laboratory settings most where factors that affect data quality are less controllable. In addition, because of our small sample size, we were unable to test for the influence of individual factors, such as eye color and glasses, on data quality. Eye trackers may be light or dark pupil-based systems, meaning that the pattern of illumination emitted by the eye tracker produces either a bright or dark image of the pupil. The type of eye tracker in conjunction with company-specific algorithms may interact with these factors to affect data quality (Holmqvist et al., 2012). Gazepoint offers a gain sweep mechanism that can be toggled on or off to assist with tracking of blue-eyed individuals. This option was not used in the present study to limit the need for participants to make decisions.

The laboratory environment in which the study took place may not be sufficiently uncontrolled to assess the problems that might arise in non-laboratory studies. This may have led to better data quality than is to be expected outside of the laboratory, and future research is needed to evaluate eye-tracking data quality outside the laboratory. Relatedly, participants in our study were undergraduate and graduate students, limiting the generalizability of our results. It may be that students in university settings are more accustomed to using computer technology in a research setting, and this positively affected data quality measures. Future research is needed to examine data quality using a more representative sample. Future research should also include a qualitative assessment of participant experience. Qualitative assessment of device use may provide information to improve protocol and training procedures. Participants in the present work were not asked to report on their experiences, so we are unable to report their perceptions. Finally, it should be noted that because data was collected from two expert users, the generalizability of the patterns observed in the baseline data is limited. It is expected that there is heterogeneity in data quality among eye-tracking experts because of differences unrelated to proper eye-tracking use or setup (e.g., eye physiology). However, although the small baseline sample may not be representative

of the variability that exists, it is encouraging that our participants obtained data quality measures that were comparable to two trained eye-tracking users.

Conclusion

An eye-tracking setup consisting of the Gazepoint GP3 controlled with PyGaze software may provide robust data collection in studies that require independent eye-tracker use. This may be most beneficial for researchers who wish to conduct studies outside the laboratory. However, given the heterogeneity of non-laboratory environments, it is recommended that researchers first assess data quality in the experimental conditions under which the data will be collected.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

- Bottos, S., & Balasingam, B. (2019). An Approach to Track Reading Progression Using Eye-Gaze Fixation Points. *CoRR*, abs/1902.03322. Retrieved from <http://arxiv.org/abs/1902.03322>
- Brand, J., Masterson, T. D., Emond, J. A., Lansigan, R., & Gilbert-Diamond, D. (2020). Measuring attentional bias to food cues in young children using a visual search task: An eye-tracking study. *Appetite*, 148, 104610. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0195666319312954>
- Calado, J., Marcelino-Jesus, E., Luis-Ferreira, F., & Sarraipa, J. (2019). *EYE-TRACKING STUDENT'S BEHAVIOUR FOR E-LEARNING IMPROVEMENT*. Retrieved from <https://doi.org/10.21125/edulearn.2019.2221>
- Čegovnik, T., Stojmenova, K., Jakus, G., & Sodnik, J. (2018). An analysis of the suitability of a low-cost eye tracker for assessing the cognitive load of drivers. *Applied Ergonomics*, 68, 1–11. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0003687017302326>
- Chen, N., Clarke, P., Watson, T., MacLeod, C., & Guastella, A. (2014). Attentional bias modification facilitates attentional control mechanisms: Evidence from eye tracking. *Biological Psychology*, 104. Retrieved from <https://doi.org/10.1016/j.biopsycho.2014.12.002>
- Coyne, J., & Sibley, C. (2016). Investigating the Use of Two Low Cost Eye Tracking Systems for Detecting Pupillary Response to Changes in Mental Workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1), 37–41. Retrieved from <https://doi.org/10.1177/1541931213601009>
- Dalmaijer, E. (2014). Is the low-cost EyeTribe eye tracker any good for research?. Retrieved from <https://doi.org/10.7287/peerj.preprints.585v1>
- Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, 46(4), 913–921.
- Duc, A. H., Bays, P., Husain, M., Kennard, C., & Leigh, R. J. (2008). Chapter 5.5 - Eye movements as a probe of attention (Vol. 171, pp. 403–411). Retrieved from <http://www.sciencedirect.com/science/article/pii/S0079612308006596>

- Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, 43(9), 1035–1045. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0042698903000841>
- Ferrari, G. R. A., Möbius, M., van Opdorp, A., Becker, E. S., & Rinck, M. (2016). Can't Look Away: An Eye-Tracking Based Attentional Disengagement Training for Depression. *Cognitive Therapy and Research*, 40(5), 672–686. Retrieved from <https://doi.org/10.1007/s10608-016-9766-0>
- Field, M., Munafo, M. R., & Franken, I. H. A. (2009). A meta-analytic investigation of the relationship between attentional bias and subjective craving in substance abuse. *Psychological Bulletin*, 135(4), 589–607. Retrieved from <https://doi.org/10.1037/a0015843>
- Funke, G., Greenlee, E., Carter, M., Dukes, A., Brown, R., & Menke, L. (2016). Which Eye Tracker Is Right for Your Research? Performance Evaluation of Several Cost Variant Eye Trackers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60, 1240–1244. Retrieved from <https://doi.org/10.1177/1541931213601289>
- Gavas, R. D., Roy, S., Chatterjee, D., Tripathy, S. R., Chakravarty, K., & Sinha, A. (2018). Enhancing the usability of low-cost eye trackers for rehabilitation applications. *PLOS ONE*, 13(6), e0196348. Retrieved from <https://doi.org/10.1371/journal.pone.0196348>
- Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2010). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: What it is and how to measure it. *Eye Tracking Research and Applications Symposium (ETRA)*. Retrieved from <https://doi.org/10.1145/2168556.2168563>
- Irwin, D. E. (1992). Visual Memory Within and Across Fixations. In ["Keith Rayner"] (Ed.) (pp. 146–165). New York, NY: Springer New York. Retrieved from https://doi.org/10.1007/978-1-4612-2852-3_9
- Kamienkowski, J. E., Ison, M. J., Quiroga, R. Q., & Sigman, M. (2012). Fixation-related potentials in visual search: A combined EEG and eye tracking study. *Journal of Vision*, 12(7), 4–4. Retrieved from <https://doi.org/10.1167/12.7.4>
- Kim, S., Pollanen, M., Reynolds, M. G., & Burr, W. S. (2020). Problem Solving as a Path to Comprehension. *Mathematics in Computer Science*. Retrieved from <https://doi.org/10.1007/s11786-020-00457-1>
- Konkel, L. (2015). Racial and Ethnic Disparities in Research Studies: The Challenge of Creating More Diverse Cohorts. *Environmental Health Perspectives*, 123(12), A297–A302. Retrieved from <https://doi.org/10.1289/ehp.123-a297>
- Kruijne, W., & Meeter, M. (2016). Implicit short- and long-term memory direct our gaze in visual search. *Attention, Perception, & Psychophysics*, 78(3), 761–773. Retrieved from <https://doi.org/10.3758/s13414-015-1021-3>
- Ladouce, S., Donaldson, D. I., Dudchenko, P. A., & Ietswaart, M. (2017). Understanding Minds in Real-World Environments: Toward a Mobile Cognition Approach. *Frontiers in Human Neuroscience*, 10, 694. Retrieved from 10.3389/fnhum.2016.00694
- Larson, M. D., & Behrends, M. (2015). Portable Infrared Pupillometry: A Review. *Anesthesia & Analgesia*, 120(6). Retrieved from https://journals.lww.com/anesthesia-analgesia/Fulltext/2015/06000/Portable_Infrared_Pupillometry__A_Review.14.aspx
- Manor, B. R., & Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of Neuroscience Methods*, 128(1), 85–93. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0165027003000151>
- Mantiuk, R., Kowalik, M., Nowosielski, A., & Bazyluk, B. (2012). Do-It-Yourself Eye Tracker: Low-Cost Pupil-Based Eye Tracker for Computer Graphics Applications. In ["Klaus Schoeffmann", "Bernard Merialdo", "Alexander G. Hauptmann", "Chong-Wah Ngo", "Yiannis Andreopoulos", & "Christian Breiteneder"] (Eds.) (pp. 115–125). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324.
- Notaro, G. M., & Diamond, S. G. (2018). Development and Demonstration of an Integrated EEG, Eye-Tracking, and Behavioral Data Acquisition System to Assess Online Learning. In *Proceedings of the 10th International Conference on Education Technology and Computers* (pp. 105–111). New York, NY, USA: Association for Computing Machinery. Retrieved from <https://doi.org/10.1145/3290511.3290526>
- Nyström, M., Andersson, R., Holmqvist, K., & van de Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, 45(1), 272–288. Retrieved from <https://doi.org/10.3758/s13428-012-0247-4>
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology (2006)*, 62(8), 1457–1506. Retrieved from <https://doi.org/10.1080/17470210902816461>
- Sibley, C., Foroughi, C., Brown, N., & Coyne, J. T. (2018). Low Cost Eye Tracking: Ready for Individual Differences Research? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1), 741–745. Retrieved from <https://doi.org/10.1177/1541931218621168>
- Sibley, C., Foroughi, C., Brown, N., Drollinger, S., Phillips, H., & Coyne, J. (2021). Augmenting Traditional Performance Analyses with Eye Tracking Metrics. In ["Hasan Ayaz" & "Umer Asgher"] (Eds.) (pp. 118–125). Cham: Springer International Publishing.
- Sibley, C., Foroughi, C., Olson, T., Moclair, C., & Coyne, J. (2017). *Practical Considerations for Low-Cost Eye Tracking: An Analysis of Data Loss and Presentation of a Solution*. Retrieved from https://doi.org/10.1007/978-3-319-58628-1_9
- Taylor, P., Bilgrien, N., He, Z., & Siegelmann, H. T. (2015). EyeFrame: Real-Time Memory Aid Improves Human Multitasking via Domain-General Eye Tracking Procedures. *Frontiers in ICT*, 2, 17. Retrieved from <https://doi.org/10.3389/fict.2015.00017>
- Zugal, S., & Pinggera, J. (2014). Low-Cost Eye-Trackers: Useful for Information Systems Research? In ["Lazaros Iliadis", "Michael Papazoglou", & "Klaus Pohl"] (Eds.) (pp. 159–170). Cham: Springer International Publishing.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.