



OPEN

## Classification of short and long term mild traumatic brain injury using computerized eye tracking

Alice Cade<sup>1,2</sup>✉ & Philip R. K. Turnbull<sup>1</sup>

Accurate, and objective diagnosis of brain injury remains challenging. This study evaluated useability and reliability of computerized eye-tracker assessments (CEAs) designed to assess oculomotor function, visual attention/processing, and selective attention in recent mild traumatic brain injury (mTBI), persistent post-concussion syndrome (PPCS), and controls. Tests included egocentric localisation, fixation-stability, smooth-pursuit, saccades, Stroop, and the vestibulo-ocular reflex (VOR). Thirty-five healthy adults performed the CEA battery twice to assess useability and test-retest reliability. In separate experiments, CEA data from 55 healthy, 20 mTBI, and 40 PPCS adults were used to train a machine learning model to categorize participants into control, mTBI, or PPCS classes. Intraclass correlation coefficients demonstrated moderate (ICC > .50) to excellent (ICC > .98) reliability ( $p < .05$ ) and satisfactory CEA compliance. Machine learning modelling categorizing participants into groups of control, mTBI, and PPCS performed reasonably (balanced accuracy control: 0.83, mTBI: 0.66, and PPCS: 0.76, AUC-ROC: 0.82). Key outcomes were the VOR (gaze stability), fixation (vertical error), and pursuit (total error, vertical gain, and number of saccades). The CEA battery was reliable and able to differentiate healthy, mTBI, and PPCS patients reasonably well. While promising, the diagnostic model accuracy should be improved with a larger training dataset before use in clinical environments.

**Keywords** Reproducibility of results, Post-concussion syndrome, Ocular motility disorders, Brain concussion, Machine learning

Traumatic brain injury (TBI) is the alteration of typical brain function that occurs after even minor trauma to the head<sup>1</sup>. There are a range of etiologies and variable presentations of mTBI, whose diagnosis is usually clinical, despite debate around the definition of mTBI<sup>2–4</sup>. Understandably, there is interest in developing biomarkers or automated objective assessments for assisting with diagnosis and tracking recovery<sup>5,6</sup>. While questionnaires<sup>7</sup> and blood biomarkers<sup>8</sup> show promise for TBI detection and tracking, biomarkers are not commercially available, and questionnaires are subjective and can lack sensitivity especially for milder presentations<sup>8,9</sup>. For example, the eye tracking-based King-Devick test has high sensitivity (86–98%) and specificity (90–96%) for moderate and severe TBIs<sup>10,11</sup>, but is much poorer at identifying the more common mild TBI (mTBI)<sup>9</sup>. This is problematic as many as 15% of people with mTBI have symptoms that last longer than the normal three-month recovery time, termed persistent post-concussion syndrome (PPCS)<sup>12</sup>. Those who progress to PPCS are thought to have several factors predisposing them to long-term symptoms such as anxiety, depression, lower socioeconomic status, reduced coping ability, pain, and motivation<sup>12</sup>. While some predisposing factors to PPCS implicate a psychological origin, the symptoms of PPCS are measurable and quantifiable, and include problems with response inhibition or executive control, visual attention, and visuospatial processing<sup>13</sup>. Those with PPCS can exhibit impaired oculomotor function when compared to those with a similar injury severity but who experience a more rapid recovery<sup>14</sup>. Being able to predict those who are at risk of PPCS would be hugely beneficial for health care systems, but first markers need to be identified.

Post-mTBI oculomotor, attentional, and exteroceptive defects can occur, but one of the most common complaints post-mTBI is that of visual disturbances, with prevalence rates between 30 and 90%<sup>15,16</sup>. Eye movements are neurologically complex actions, involving many areas of the brain, so unsurprisingly functional visual defects are a common finding following all severities of TBI<sup>15,17</sup>. Visual defects can result from damage to the primary visual pathway to the occipital lobe, along the parieto-occipital pathway, frontal lobes, cerebellum, or to the oculomotor cranial nerves themselves<sup>18</sup>. Furthermore, eye movements also require integrated action from the

<sup>1</sup>School of Optometry and Vision Science, The University of Auckland, Private Bag 92019, Auckland 1023, New Zealand. <sup>2</sup>New Zealand College of Chiropractic, Auckland, New Zealand. ✉email: a.cade@auckland.ac.nz

proprioceptive<sup>19</sup>, attentional<sup>20</sup>, visual salience<sup>21</sup>, and vestibular<sup>22</sup> brain centers, so injury to any of these areas can also lead to visual abnormalities<sup>21,23</sup>. Eye tracking has been used to highlight many defects in those with mTBI, including poorer fixation<sup>24</sup>, increased gaze-error during pursuit movements<sup>25</sup>, increased saccade latency<sup>26,27</sup>, reduced vestibulo-ocular reflex (VOR) gain<sup>28</sup>, poorer performance on the Stroop test<sup>29</sup>, and higher egocentric localisation errors<sup>30,31</sup>, but no one test is diagnostic. Therefore, a comprehensive battery of computerized eye tracking assessments (CEAs) would seem to be ideally positioned to objectively detect mTBI after an injury<sup>32</sup>. Previous researchers have created eye tracking batteries that are sensitive and specific to identifying mTBI<sup>33</sup>, but they require restrained participants or skilled clinicians to operate and are either expensive or no longer unavailable<sup>34–36</sup>. While using CEAs to identify those with mTBI is not a novel concept, disentangling the differences between healthy people, mTBIs, and those with PPCS is more complex. This study aimed to assess the feasibility of developing a machine learning model that used the outputs from a battery of CEAs to firstly differentiate those with mTBI from controls, and then attempt to identify differences in CEA results between recent mTBIs (symptomatic, but under three months since injury) and those with PPCS (mTBI symptoms present for over three months). This project is the first step towards building normative and brain-injured data sets from which accurate diagnoses can be made, as well as close measurements of injury change over time.

## Methods

### Trial design and participants

Baseline data from 55 control, 20 mTBI, and 40 PPCS participants were pooled from baseline measures across three other studies that used the same CEA battery and test regime. The first assessed the test–retest reliability of the CEA tests in healthy people, the second assessed the effects of a vibratory intervention on the cervical spine in controls and mTBI, and the third was a randomised controlled interventional trial for people with PPCS<sup>37</sup>. All data used for this study were pre-intervention (i.e., baseline) measures from participants. The control group had no history of TBI, self-reported normal vision, and no known oculomotor deficits. All participants with mTBI or PPCS were diagnosed by their primary health care provider prior to enrolment, had self-reported normal vision (other than symptoms that started after the brain injury), and were able to understand the study information and consent processes. Additionally, a subset of the healthy control group had the test battery repeated (with a shuffled test order) in the same session, to evaluate CEA test reliability.

Participants were free to withdraw from the studies at any stage and gave written informed consent. All experimental protocols and procedures were approved by the New Zealand Health and Disability Ethics Committee (HDEC 19/CEN/130) and the studies complied with the Declaration of Helsinki.

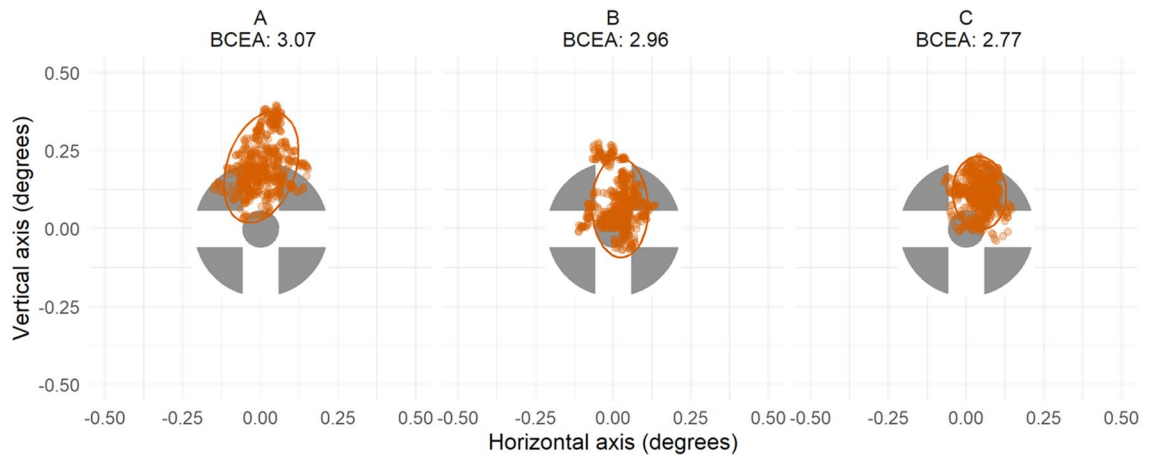
### Materials/computer set-up

A laptop mounted eye tracker (Tobii 5, Tobii Group, Stockholm, Sweden, see Supplementary Materials) was used to record binocular gaze (133 Hz) and head position data (33 Hz) samples per second. Visual stimuli were presented on a laptop computer (Surface Book 2, Microsoft, Redmond, USA), and the eye tracker was calibrated using the Tobii calibration software for each participant prior to running the CEAs. During testing, the participant sat approximately 70 cm from the screen, although the CEAs accounted for changes in angular extent due to changes in viewing distance. The target stimuli consisted of a black cross-shaped target presented on a white background<sup>21</sup>. The tracker used did have a lower sampling rate than what is generally considered the minimum (240 Hz) for saccades<sup>38</sup>. Lower sampling rates can lead to aliasing of peak saccade velocity, affecting results. Similarly, over filtering the data can lower peak velocity, especially in data recordings with lower sampling rates<sup>39</sup>, hence why a filter was not used on these data. While some research suggests that good approximations of peak velocity can be obtained when modelling data from sampling rates as low as 50 Hz<sup>40</sup>, the decision was made to use multiple measures to investigate saccades, such as correctly performing a test or decision-making latency, rather than to solely rely on peak saccade velocity as an outcome.

### Computerized eye tracking assessments

Six tests with previously established diagnostic value in differentiating mTBI<sup>32,37</sup>, were used in this study.

1. *Egocentric localisation*. Participants were asked to move to align the center of their head with the target that appeared in a random screen location, ten times. The main outcomes were mean offset error and the mean trial completion time. The target was confined to the screen centre, with 15% screen width padding across the top, left, and right, and 30% padding from the bottom of the screen.
2. *Fixation stability*. Participants were instructed to maintain stable fixation on a centrally placed target for three trials of 10 s each. The main outcomes were the 95% bivariate contour ellipse area of fixation (BCEA,  $\log_{10} \text{minarc}^2$ )<sup>41</sup>, and a vector of mean gaze error, Fig. 1.
3. *Smooth pursuit*. Participants gaze followed a moving target as it traversed the screen in a Lissajous pattern, moving sinusoidally with a width of 11° and height of 6.5° at 10°/s for four trials of 30 s each. The outcomes measures were mean gaze offset error, total gain, and the number of catch-up saccades.
4. *Saccade test*. 14 pro-saccade and 14 anti-saccade tasks were randomly interleaved and assessed in one continuous run. In the saccade task the target appeared centrally on the screen for a random interval between 0.5–1.0 s. After this time the target color would change color to either green (for the pro-saccade task) or red (for the anti-saccade task) and stay visible for a further 0.5 s before disappearing. While the central target was still visible, a black secondary target, otherwise identical in size and shape to the main target, would appear at three degrees left or right of the screen for 0.2 s. Main outcomes were saccade latency—the time taken to decide where to move gaze after target presentation—and the number of correctly performed trials—did the participant look towards (pro-saccade task) or away (anti-saccade task) from the secondary target.

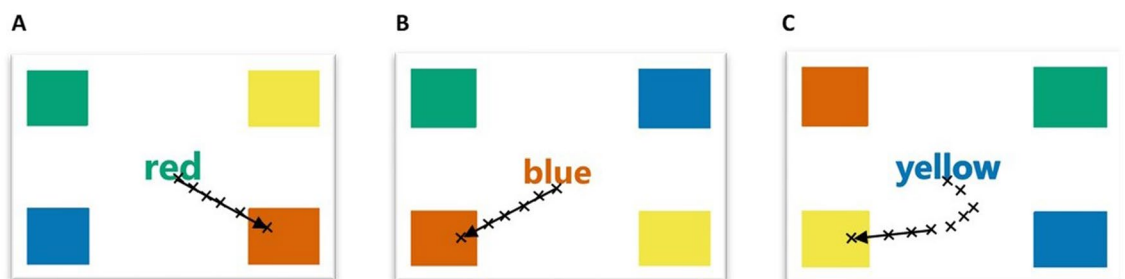


**Figure 1.** An example raw data from three trials of the fixation stability test. The orange points represent the participant's gaze samples on the target (scaled size and opacity) when it was on the screen. The 95% bivariate contour ellipse area (BCEA, orange ellipse) is listed above each plot.

5. *The Stroop tests.* Adapted from the Stroop Color-Word test<sup>29</sup>. Part one involved directing gaze towards the color indicated by the stimulus word for 15 trials, then part two required participants to respond based on the stimulus font color, for another 15 trials (Fig. 2). Stroop outcomes included mean saccade decision making latency, total trial time, and proportion of correct trials. Prior to the test, participants viewed the 4 corner color blocks to familiarize themselves with the screen set-up. Each trial began with black words on a white screen of “Look here to start”, requiring active gaze for the trial to begin. For both parts a 50-point Arial font stimulus word appeared in the center of the screen. The stimulus words presented were “red,” “blue,” “green,” and “yellow,” randomly selected. The font color of each word was presented in red, blue, green, or yellow, however the word could not match the color of the font that the word was written in. For example, for the word red, the only possible font color permutations were green, blue, or yellow. The font colors were based on landmark colors that have unique psychophysiological bases, meaning they are easily recognized<sup>42</sup>. The colors used were adapted to color-blindness safe colors of blue (RGB: 0, 114, 178), red (RGB: 213, 94, 0), green (RGB: 0, 158, 115), and yellow (RGB: 240, 228, 66)<sup>43</sup>, and extended close enough to fixation that they did not need foveation to identify.
6. *The vestibulo-ocular reflex test.* The VOR was assessed with an active head impulse test<sup>44</sup> by having the target appear centrally on the screen for 1.5 s, after this time the target turned green. The instructions given to each participant were to “Keep your eyes fixed on the central target as closely as possible then after it turns green, quickly turn your head to the left or right while keeping your eyes fixed on the central target.” There were 20 trials, 10 where participants turned their head to the left, and 10 to the right. Participants were free to choose which direction they first turned their head, then directed to alternate their head rotation in each subsequent trial<sup>45</sup>. Gaze was recorded for the first 15 degrees of head rotation. The main outcomes were fixation stability (as 95% BCEA), head to eye velocity gain, and the number saccades during each trial.

### Eye gaze data processing

Eye gaze data was captured continuously throughout the trials, and any trials where the participant moved out of the nominal measurement range of the eye tracker during the trial (e.g., closer than 500 mm or farther than



**Figure 2.** An example of the data from part one of the Stroop test, with X denoting gaze samples. Panel (A) shows gaze data from a single correctly performed part one trial, and panel (B) shows a correctly performed part two trial. Panel (C) shows a correctly performed part one trial is shown where the participant has corrected their gaze path partway through a trial.

950 mm) had those trials excluded (only 4 trials were excluded throughout all data collection)<sup>46,47</sup>. Any trials with a saccade latency below 60 ms—5 trials total—were removed as the participants gaze was likely already moving, much like a false start in a running race, with an absolute minimum latency being previously reported at around 80 ms<sup>48</sup>. Gaze data, in screen coordinates, was converted to degrees of visual angle and gaze velocity was calculated as degrees of visual angle per second<sup>49</sup>. Blinks resulting in loss of signal were removed before analysis. As the Tobii 5 eye tracker has a sampling rate of 133 Hz, saccades were identified by time (at least 50 ms) and velocity (over 100 degrees/sec) to avoid misidentification of a saccade by velocity alone<sup>50</sup>.

### Statistical analyses

Statistical analysis was performed using R software (version 4.0.3, <https://www.r-project.org/>) in RStudio (version 1.3.1093, <https://posit.co/>) using the `stats`, `irr`, and `caret` libraries. Bland–Altman plots with mean differences and limits of agreement<sup>51</sup> and two-way intraclass correlation coefficients (ICCs) for consistency were used to assess reliability of CEA tests<sup>52</sup>. ICC values were described as excellent when greater than 0.90, good when between 0.75 and 0.90, moderate from 0.50 to 0.75, and poor when below 0.50<sup>53–55</sup>. For machine learning analysis, CEA outcome data were first normalized and centered, then with a common random seed, divided into a 70% training set and 30% test set. Each of the models were trained using 10 × cross-validation to prevent overfitting, using kappa, which compared to accuracy, accounts for the relative weightings of each class. The performance of each model and the relative importance of each variable was stored. To prevent chance overfitting to a specific random seed, the entire process was repeated using 10 different random seeds, which also enabled calculation of confidence intervals for model performance. To calculate ROC characteristics, pairwise ROC curves were computed with overall performance taken as a mean of all comparisons, Fig. 3.

Differences in CEA parameters between the three groups were compared using non-parametric Kruskal–Wallis test, statistical significance for all tests was set at  $p < 0.05$  and adjusted using Holm multiplicity correction as appropriate<sup>56</sup>.

### Results

One hundred and fifteen age-balanced (control:  $30.1 \pm 7.9$ , mTBI:  $25.5 \pm 2.8$ , PPCS:  $25.8 \pm 4.3$ ,  $\chi^2_{(1, N=115)} = 4.94$ ,  $p = 0.085$ ) and gender-balanced (female control: 65%, mTBI: 55%, PPCS: 70%,  $\chi^2_{(1, N=115)} = 3.53$ ,  $p = 0.171$ ) participants completed the whole experiment, with one mTBI participant withdrawing as the computer screen caused them painful photophobia. Their data was not included in the analysis.

### Repeatability of CEA measures

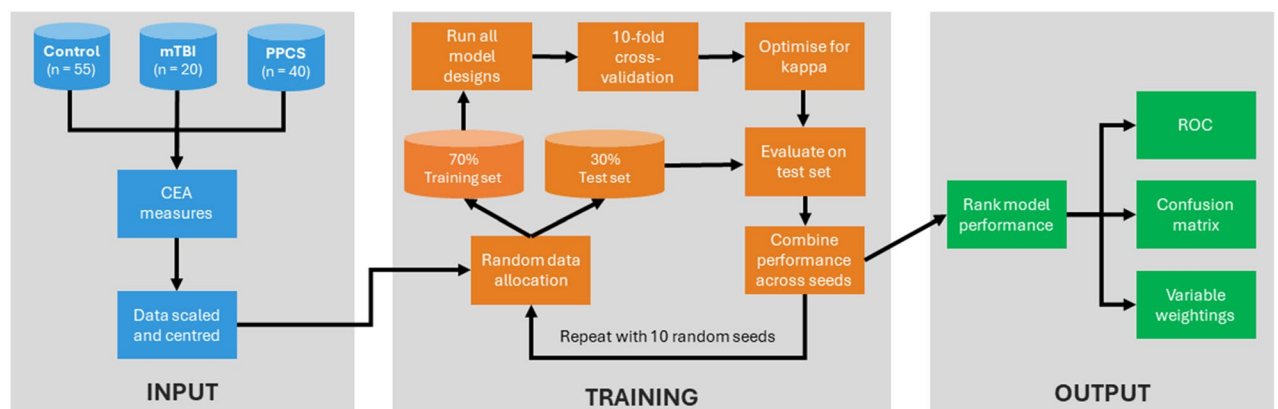
For ease of interpretation of the large number of test outcomes, all test results are provided in a Supplementary Table, and all outcomes with ICC values over 0.50 are shown in Fig. 4. Discussion of post-hoc comparisons is provided in the text below.

Analysis of gaze stability (BCEA and offset error) showed excellent reliability in fixation (ICC: 0.90–0.98) and VOR (ICC: 0.87–0.96). VOR gain reliability was also excellent (ICC: 0.94), but VOR time to maximum gain was poor (ICC: 0.24). Pursuit and egocentric localisation showed moderate to excellent reliability (ICC: 0.72–0.97).

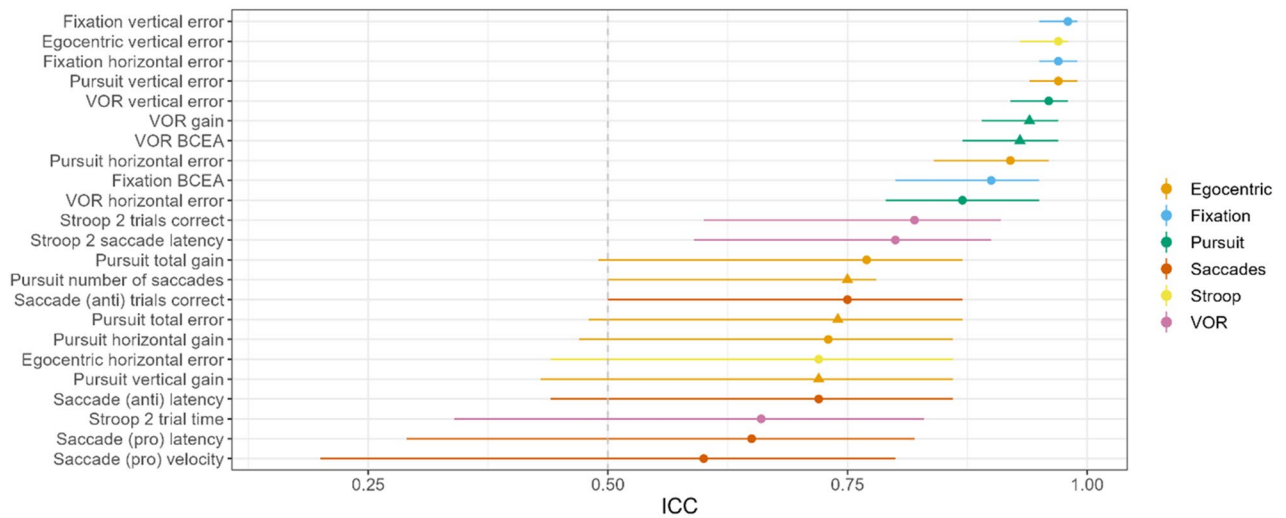
The ICC for Stroop and saccade parameters was moderate for pro-saccade latency (0.65), velocity (0.60), anti-saccade latency (0.72) and anti-saccade correct trials (0.75), but poorer for pro-saccade correct trials (0.46) and anti-saccades velocity (0.46). Overall reliability in the Stroop test was higher for part two (ICC: 0.66–0.82), compared to part one (ICC:  $-0.04$ –0.48). There was also some evidence of a learning effect during Stroop testing, with the probability of performing a trial correctly increasing with each trial for both part one ( $\chi^2(1, N = 35) = 0.72$ ,  $p = 0.006$ ) and two ( $\chi^2(1, N = 35) = 30.12$ ,  $p < 0.001$ ).

### Machine learning analyses

Machine learning modelling was undertaken using a range of models to categorize participants into groups of control, mTBI, and PPCS. The most performant model used extreme gradient boosting with dropout (xgbDART

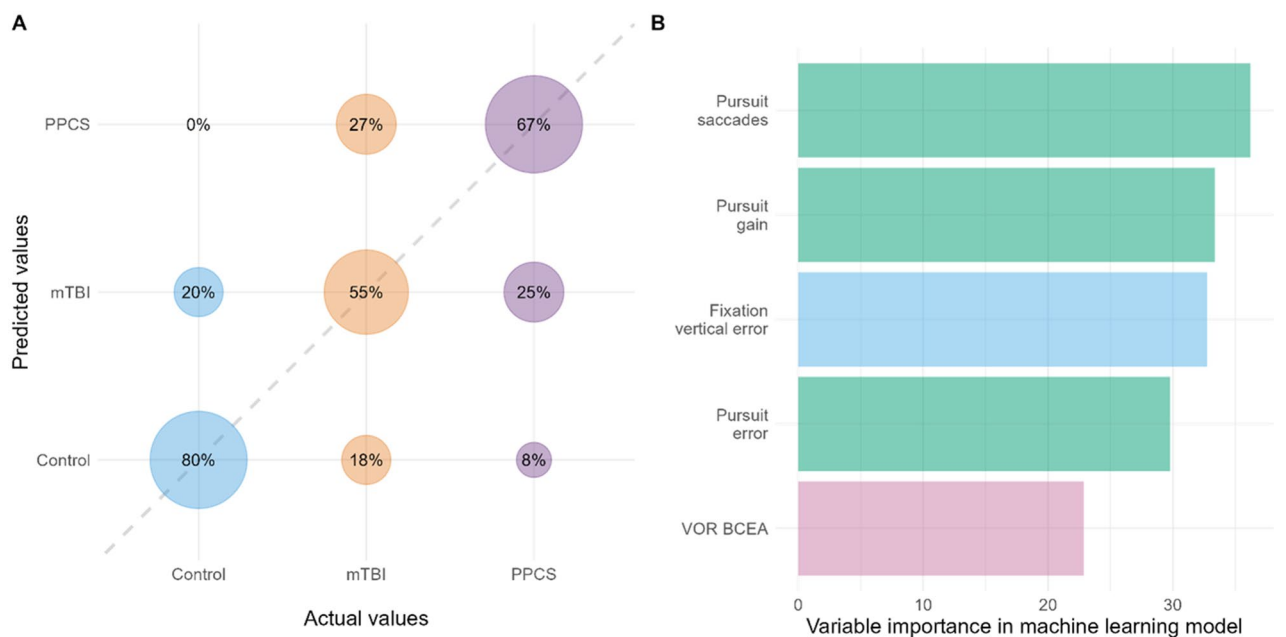


**Figure 3.** A schematic diagram showing the steps involved with training and evaluating the machine learning models.



**Figure 4.** Intraclass correlation coefficients (ICC, points, with 95% confidence intervals) for all test outcomes with an ICC over 0.50. The outcomes that were ranked in the top five variables for the machine learning modelling have a triangle marker.

), which yielded a mean kappa of  $0.75 \pm 0.20$  and a mean 3-class pairwise AUC-ROC of 0.82 over the 10 random seeds. The balanced accuracy (mean of sensitivity and specificity) of the control, mTBI, and PPCS groups was 0.83, 0.66, and 0.76, respectively (Fig. 5A). This means that most brain-injured participants were correctly categorized as to their classification of injury (specificity mTBI: 0.77, PPCS: 0.88), and suggests the model was able to glean meaningful differences in the CEA data between groups. Key outcomes for modelling were the VOR (gaze stability), fixation (vertical error), and pursuit (total error, vertical gain, and number of saccades, Fig. 5B). While most CEA parameters had reasonable reliability, there was no correlation between each tests' ICC value and the importance of their weighting in machine learning modelling (Kendall's correlation:  $\tau_b = 0.25, p = 0.08$ ).



**Figure 5.** (A) Machine learning analysis confusion matrix for predicting control, mTBI, or PPCS. Overall, the balanced accuracy of 0.83, 0.66, and 0.76 for controls, mTBI, and PPCS, respectively. Further, the most common error in classifying mTBI was PPCS, and for PPCS the most common error was labelling as mTBI. As the distinction between mTBI and PPCS are arbitrarily defined as symptoms beyond three months, it is perhaps not surprising to see overlap between their symptomologies. (B) shows the top 5 outcomes that were most important to classifying groups in the machine learning model. There was no significant correlation between ICC values for outcomes and the importance of their weighting in machine learning modelling, Kendall's correlation:  $\tau_b = [0.25], p = .08$ .

### Differences in CEA outcomes between groups

Gaze stability in VOR was different ( $H_{(2)} = 261.68, p < 0.001$ ) between healthy and brain-injured participants ( $p < 0.001$ , control:  $3.26 \pm 0.94$ , mTBI:  $3.94 \pm 0.01$ , PPCS:  $3.98 \pm 0.90$ ), but not between mTBI and PPCS ( $p = 0.930$  mTBI:  $3.94 \pm 0.01$ , PPCS:  $3.98 \pm 0.90$ ). Gaze error during pursuits was different between all groups ( $H_{(2)} = 73.16, p < 0.001$ ) with mTBI ( $0.30 \pm 0.23^\circ$ ) being significantly worse compared to controls ( $0.003 \pm 0.26^\circ, p < 0.001$ ), whereas PPCS errors (PPCS:  $0.15 \pm 0.34^\circ$ ) were roughly midway between the mTBI and control means ( $p < 0.001$ ). Vertical pursuit gain ( $H_{(2)} = 64.53, p < 0.001$ ) was different between groups, with mTBIs having the highest gain ( $1.04 \pm 0.03$ ) compared to controls ( $1.03 \pm 0.03, p = 0.008$ ), while those with PPCS had a vertical pursuit gain closest to the ideal of 1.00 ( $1.02 \pm 0.04$ ). Also during pursuit testing, the number of catch-up saccades during a trial varied between groups ( $H_{(2)} = 21.35, p < 0.001$ ), with the mean number saccades per trial being  $0.50 \pm 1.00, 0.38 \pm 0.70$ , and  $1.05 \pm 1.61$ , for control, mTBI and PPCS, respectively. The number of saccades was not different between mTBI and controls ( $p = 0.470$ ) but was between controls and mTBI ( $p = 0.001$ ) and controls and PPCS ( $p < 0.001$ ). For vertical offset error in fixation ( $H_{(2)} = 67.08, p < 0.001$ ), only the PPCS group was significantly different ( $0.32 \pm 0.75^\circ$ ), with their gaze sitting farther above the target compared to other groups (control:  $-0.29 \pm 0.47^\circ$ , mTBI:  $-0.38 \pm 0.80^\circ, p < 0.001$ ). All significant and non-significant differences between groups for all outcomes are shown in the Supplementary Table.

## Discussion

### Key findings

This study found moderate to excellent test–retest reliability for a range of CEA measures, that also had good diagnostic value for differentiating control, mTBI and PPCS cohorts. Our results allowed a machine learning model can be trained with performance equivalent or better than previous eye tracking test batteries for mTBI<sup>33–35</sup>, even with a modest sample size. The CEA tests used in this study were specifically designed to be simple to use and to not require complex hardware or a highly trained clinician to administer.

### Most informative tests outcomes

Both the VOR and pursuit test had test parameters that were important in differentiating between the three classes, and the parameters that were most informative for the modelling tended to be different between classes using frequentist statistics. As the machine learning model also looks for complex relationships between variables, at least with our dataset it seems that such correlative measures or interactions did not add significant predictive value to the model. This may mean that the size of the model could be reduced with minimal impact on accuracy, allowing it to run on a wider range of hardware.

In terms of why VOR and pursuit may be most informative, it is important to note that in this study participants' heads were not restrained. This permitted more naturalistic head movement and vestibular input during testing, and both VOR and pursuit require more complex cognitive integrations (including from VOR) before the resultant eye movement output. Proprioceptive drive from head rotation in the VOR test impacts gaze stability via cerebellar-vestibular pathways, whereas pursuit movements are mediated by predictive cerebellar-frontal pathways<sup>57,58</sup>. Perhaps, it is the combination of these two pathways—vestibular and frontal—for gaze stability and pursuit that uncover brain-injury deficits more effectively compared to simpler tests, such as saccades, which does not include as much proprioceptive input or prediction of movement. The VOR test's importance may also further implicate the role of the injured cervical spine in mTBI symptomatology due to their similar mechanism of injury<sup>59</sup>. Cervical afferents, especially those of the upper cervical spine, heavily influence vestibulo-cerebellar components of gaze stabilization pathways and provide a secondary estimate of head movement, augmenting VOR function<sup>60</sup>. Disordered cervical proprioception post-injury may have affected brainstem activation or vestibulo-cerebellar components of the VOR pathway impacting its function<sup>57,58,61</sup>.

During pursuit movements, neurons in the vestibular nuclei receive direct cerebellar floccular input and provide premotor input to extraocular muscles helping to coordinate eye and head movement<sup>62</sup>. Due to the latency in the visual system, accurate pursuit requires anticipating target movement and learning target patterns, which originate from the middle temporal area and medial superior temporal area, and receive inputs from the frontal eye fields<sup>63</sup>. These areas, collectively, can be hypo-perfused and exhibit reduced glucose metabolism post-mTBI<sup>64</sup>, which is suggestive of poorer function and potentially explains the importance of pursuit gaze error in identifying mTBI. Poor gaze error would then lead to a need to catch the gaze up to the target with saccades, as was found in those with PPCS, further augmenting pursuit in mTBI classification and differentiating chronic symptomatology from recent injury.

During fixation testing, vertical error was of importance, which may also implicate cervical spine injury. Previous research showed that vibration to the cervical spine in mTBI worsened vertical fixation error, but a manual therapy aimed at reducing proprioceptive dysfunction in the spine after head injury improved vertical error<sup>37</sup>. These data suggest that just augmenting proprioceptive drive from the post-head-injured spine is not the key for improving fixation, but that reducing disordered proprioception is crucial.

Further information can also be gleaned from the errors that the model made, of which the most common was a misclassification between mTBI and PPCS. As the distinction between these two groups is an arbitrary period, the model may have identified participants that were heading towards chronicity of symptoms, and PPCS participants may have been returning to normality. It is possible that some outcomes tested may also help differentiating mTBI from PPCS specifically, such as the gaze error during pursuit testing, which was poorer in mTBI and better in PPCS, or the number of saccades in pursuits, which was significantly poorer in PPCS compared to the other groups. These data may guide further research into differentiating chronic from recent brain-injury, but more work is required as mechanism of injury or predisposing factors could also explain these differences.

### Performing trials correctly

Our pro-saccade latency (ICC = 0.59) and anti-saccades latency (0.72) ICC values were similar to previous findings (pro: = 0.66–0.79<sup>65–67</sup>, anti: = 0.64–0.73<sup>65–68</sup>). The number of trials performed incorrectly showed lower reliability for pro-saccades than anti-saccades, other research shows similar trends of pro-saccade ICCs of error rates being much lower (0.14<sup>65</sup>) compared to anti-saccade error rates (0.79–0.91<sup>65,67</sup>). Some of these differences may be explained by whether the pro- and anti-saccade tasks were interleaved, which has been shown to increase error rates in healthy controls compared to blocked-tasks error rates<sup>69</sup>. This difficulty in switching between tasks may also have explained the markedly longer saccade latencies and smaller difference between pro- and anti-saccade latency seen in this study, compared to previous research<sup>26</sup>. Switching between dual-tasks inhibits excitatory drive from the frontal eye fields and inhibit reflexive pro-saccades—essentially making an anti-saccade easier to perform and reducing the difference between pro- and anti-saccade latency<sup>70</sup>. For the Stroop test, part two error was consistent with previous research<sup>71,72</sup>, but our Part one error rates were lower (~15%) than normative data (~30%)<sup>71,72</sup>, but with poorer reliability (ICC = -0.04) than reported by others<sup>73,74</sup>. Further investigation shows this to be confusion between the green and yellow colors—potentially explaining why there appeared to be a learning effect during Stroop testing—which was corrected through a more comprehensive introduction<sup>37</sup>.

### Future work

Regarding future use of this battery of CEAs, while this study focused on whether it captured differences in eye movement behaviours between mTBI and control groups, an alternate method could be to assess for within-subject differences in performance from a pre-season baseline. Used in this way, this CEA battery could track an athlete over the course of many games to assess for markers of cognitive dysfunction, which may not reach the level of even a mild TBI. Further work to understand normal longitudinal variations in CEA performance would be required, but this could greatly increase the sensitivity of the tests for detecting cognitive impairments.

### Strengths and limitation of the study

While this study piloted a new CEA test battery, a larger population will allow for more robust conclusions about which outcomes can best differentiate and predict mTBI and PPCS will make the eye tracking battery a useful tool in clinical practice. Longitudinal data collection could also be useful, to investigate which predict recovery, persist the longest, and which are the most sensitive to intervention. Longitudinally following a participant from their initial injury throughout their recovery, while tracking CEA outcomes, would create a more detailed dataset and potentially allow a more accurate model to be built. It could also help to build a metric of injury magnitude by correlating CEA outcomes with injury type, injury mechanism, symptoms experienced, and recovery time. This future modelling could help to predict an individual's progress, track their recovery, and provide objective data to inform treatment pathways.

A new metric of injury magnitude may also be required, and perhaps a method to classify their mechanism of injury. The Rivermead Post-Concussion Questionnaire<sup>75</sup> was slated to be this metric but inclusions as a covariate made no difference to any statistical modelling. Historically, differentiating mild from moderate and severe mTBI is problematic<sup>76,77</sup>, as well as being able to account for injury magnitude may help to better predict which group a participant falls into.

The eye tracker itself was also a possible limiting factor, for example, peak velocity of saccades may have been affected by the sampling rate of the eye tracker as they were at the lower end of what was expected<sup>38</sup>. Previous research has suggested sampling rates below 240 Hz can lead to aliasing—and concomitant lower peak velocities<sup>39</sup>. It is possible that future studies, using a higher sampling rate when recording data, may find some differences in saccade peak velocities compared to this study—a possibility that could be explored in further research. Aliasing may also explain why normative data for VOR gain is lower (~0.95–0.98<sup>78</sup>) than what this study found. A likely explanation could be the lower sampling rate (133 Hz) of the eye tracker compared to the 500 Hz<sup>79</sup> or 1000 Hz<sup>44</sup> sampling rates used in previous work. Potentially, the lower sampling rate led to blunting of velocity peaks altering the overall gain calculation. Additionally, the sampling rate of head tracking with the Tobii eye tracker is 33 Hz<sup>46</sup> which, as per gaze tracking, could have led to underestimating head peak velocity, further affecting gain calculations. To fully explore this possibility further study would be necessary comparing the eye tracker used in this study to one with a higher sampling rate on the same population, although this problem is likely to reduce over time with the implementation of more advanced hardware.

### Conclusions

The results from this study suggest that the battery of CEAs is reliable and can be used to reasonably accurately predict if a participant is a healthy control, has an mTBI, or has PPCS. These findings should be tempered by the fact that, while reasonable performance was obtained from our data, the diagnostic model was trained on a limited range and scope of information and larger training data sets are required before they can inform clinical practice.

### Data availability

The data that support the findings of this study are available on reasonable request from the corresponding author.

Received: 27 February 2024; Accepted: 29 May 2024

Published online: 03 June 2024

## References

1. Maas, A. I. R. *et al.* Traumatic brain injury: Integrated approaches to improve prevention, clinical care, and research. *Lancet Neurol.* **16**, 987–1048 (2017).
2. Carroll, L. J., Cassidy, J. D., Holm, L., Kraus, J. & Coronado, V. G. Methodological issues and research recommendations for mild traumatic brain injury: The WHO collaborating centre task force on mild traumatic brain injury. *J. Rehabil. Med.* **43**, 113–125. <https://doi.org/10.1080/16501960410023877> (2004).
3. Kay, T. *et al.* American Congress of Rehabilitation Medicine Mild Traumatic Brain Injury Committee of the Head Injury Interdisciplinary Special Interest Group. Definition of mild traumatic brain injury. *J. Head Trauma Rehabil.* **8**, 86–87 (1993).
4. Frieden, T. R., Houry, D. & Baldwin, G. *Report to Congress on Traumatic Brain Injury in the United States: Epidemiology and Rehabilitation.* [https://www.cdc.gov/traumaticbraininjury/pdf/TBI\\_Report\\_to\\_Congress\\_Epi\\_and\\_Rehab-a.pdf](https://www.cdc.gov/traumaticbraininjury/pdf/TBI_Report_to_Congress_Epi_and_Rehab-a.pdf) (2015).
5. Ciuffreda, K. J. & Tannen, B. Future directions in neuro-optometry. *Concussion* **5**, 80–2056. <https://doi.org/10.2217/cnc-2020-0013> (2020).
6. Bentley, M., Singhal, P., Christey, G. & Amey, J. Characteristics of patients hospitalised with traumatic brain injuries. *New Zealand Med. J.* **135**, 111–127 (2022).
7. Legarreta, A. D., Mummareddy, N., Yengo-Kahn, A. M. & Zuckerman, S. L. On-field assessment of concussion: Clinical utility of the King-Devick test. *Open Access J. Sports Med.* **10**, 115–121 (2019).
8. Bazarian, J. J. *et al.* Serum GFAP and UCH-L1 for prediction of absence of intracranial injuries on head CT (ALERT-TBI): A multicentre observational study. *Lancet Neurol.* **17**, 782–789 (2018).
9. Silverberg, N. D., Luoto, T. M., Öhman, J. & Iverson, G. L. Assessment of mild traumatic brain injury with the King-Devick Test\* in an emergency department sample. *Brain Inj* **28**, 1590–1593 (2014).
10. Hecimovich, M., King, D., Dempsey, A. R. & Murphy, M. The King-Devick test is a valid and reliable tool for assessing sport-related concussion in Australian football: A prospective cohort study. *J. Sci. Med. Sport* **21**, 1004–1007 (2018).
11. Galetta, K. M. *et al.* The King-Devick test of rapid number naming for concussion detection: Meta-analysis and systematic review of the literature. *Concussion* **10**, 1–12. <https://doi.org/10.2217/cnc.15.8> (2016).
12. Wood, R. L. Understanding the ‘miserable minority’: A diathesis-stress paradigm for post-concussional syndrome. *Brain Injury* **18**, 1135–1153. <https://doi.org/10.1080/02699050410001675906> (2004).
13. Taghdiri, F., Varriano, B. & Tartaglia, M. C. Assessment of oculomotor function in patients with postconcussion syndrome: A systematic review. *J. Head Trauma Rehabil.* **32**, E55–E67. <https://doi.org/10.1097/HTR.000000000000286> (2017).
14. Heitger, M. H. *et al.* Impaired eye movements in post-concussion syndrome indicate suboptimal brain function beyond the influence of depression, malingering or intellectual ability. *Brain* **132**, 2850–2870 (2009).
15. Kapoor, N. & Ciuffreda, K. J. Vision disturbances following traumatic brain injury. *Curr. Treat. Options Neurol.* **4**, 271–280 (2002).
16. Thiagarajan, P., Ciuffreda, K. J. & Ludlam, D. P. Vergence dysfunction in mild traumatic brain injury (mTBI): A review. *Ophthalmic Physiol. Opt.* **31**, 456–468 (2011).
17. Ciuffreda, K. J., Ludlam, D. & Thiagarajan, P. Oculomotor diagnostic protocol for the mTBI population. *Optometry* **82**, 61–63 (2011).
18. Ciuffreda, K. J., Ludlam, D. P., Yadav, N. K. & Thiagarajan, P. Traumatic brain injury: Visual consequences, diagnosis, and treatment. *Adv. Ophthalmol. Optom.* **1**, 307–333 (2016).
19. Treleaven, J. & Takasaki, H. Characteristics of visual disturbances reported by subjects with neck pain. *Man Ther.* **19**, 203–207 (2014).
20. Diwakar, M. *et al.* Filling in the gaps: Anticipatory control of eye movements in chronic mild traumatic brain injury. *Neuroimage Clin.* **8**, 210–223 (2015).
21. Lencer, R. & Trillenber, P. Neurophysiology and neuroanatomy of smooth pursuit in humans. *Brain Cogn.* **68**, 219–228 (2008).
22. Purves, D., Augustine, G., Fitzpatrick, D. & *et al.*, Editors. Central Vestibular Pathways: Eye, Head, and Body Reflexes. In *Neuroscience*, 2nd edn 2–4 (Sinauer Associates Sunderland, Massachusetts, 2001).
23. Troost, B. T. *The neurology of eye movements neurology* Vol. 34 (Oxford University Press, New York, 1984).
24. Hunfalvy, M., Murray, N. P. & Carrick, F. R. Fixation stability as a biomarker for differentiating mild traumatic brain injury from age matched controls in pediatrics. *Brain Inj.* **35**, 209–214 (2021).
25. Maruta, J. *et al.* Frequency responses to visual tracking stimuli may be affected by concussion. *Mil Med.* **182**, 120–123 (2017).
26. DiCesare, C. A., Kiefer, A. W., Nalepka, P. & Myer, G. D. Quantification and analysis of saccadic and smooth pursuit eye movements and fixations to detect oculomotor deficits. *Behav. Res. Methods* **49**, 258–266 (2017).
27. Ting, W. K. C., Schweizer, T. A., Topolovec-Vranic, J. & Cusimano, M. D. Antisaccadic eye movements are correlated with corpus callosum white matter mean diffusivity, stroop performance, and symptom burden in mild traumatic brain injury and concussion. *Front. Neurol.* **6**, 271 (2016).
28. Scherer, M. R., Shelhamer, M. J. & Schubert, M. C. Characterizing high-velocity angular vestibulo-ocular reflex function in service members post-blast exposure. *Exp. Brain Res.* **208**, 399–410 (2011).
29. Vakil, E., Weisz, H., Jedwab, L., Groswasser, Z. & Aberbuch, S. Stroop color-word task as a measure of selective attention: Efficiency in closed-head-injured patients. *J. Clin. Exp. Neuropsychol.* **17**, 335–342 (1995).
30. Chokron, S. & Bartolomeo, P. Correlation between the position of the egocentric reference and right neglect signs in left-brain-damaged patients. *Brain Cogn.* **43**, 99–104 (2000).
31. Karnath, H. O., Fetter, M. & Niemeier, M. Disentangling gravitational, environmental, and egocentric reference frames in spatial neglect. *J. Cogn. Neurosci.* **10**, 680–690 (1998).
32. Cade, A. & Turnbull, P. R. K. Clinical testing of mild traumatic brain injury using computerised eye-tracking tests. *Clin. Exp. Optom.* **105**, 680–686. <https://doi.org/10.1080/08164622.2021.2018915> (2022).
33. Hoffer, M. E. *et al.* The use of oculomotor, vestibular, and reaction time tests to assess mild traumatic brain injury (mTBI) over time. *Laryngoscope Investig. Otolaryngol.* **2**, 157–165 (2017).
34. Bin-Zahid, A. *et al.* Eye tracking as a biomarker for concussion in children. *Clin. J. Sport Med.* **30**, 433–443 (2020).
35. Samadani, U. *et al.* Eye tracking detects disconjugate eye movements associated with structural traumatic brain injury and concussion. *J. Neurotrauma* **32**, 548–556 (2015).
36. Niehorster, D. C., Cornelissen, T. H. W., Holmqvist, K., Hooge, I. T. C. & Hessels, R. S. What to expect from your remote eye-tracker when participants are unrestrained. *Behav. Res. Methods* **50**, 213–227 (2018).
37. Cade, A. Computerised Eye Tracker Assessment in Young Adults with Mild Traumatic Brain Injury. (University of Auckland, 2023).
38. Garbutt, S., Harwood, M. R. & Harris, C. M. Comparison of the main sequence of reflexive saccades and the quick phases of optokinetic nystagmus. *Br. J. Ophthalmol.* **85**, 1477–1483 (2001).
39. Mack, D. J., Belfanti, S. & Schwarz, U. The effect of sampling rate and lowpass filters on saccades: A modeling approach. *Behav. Res. Methods* **49**, 2146–2162 (2017).
40. Wierfs, R., Janssen, M. J. A. & Kingma, H. Measuring saccade peak velocity using a low-frequency sampling rate of 50 Hz. *IEEE Trans. Biomed. Eng.* **55**, 2840–2842 (2008).
41. Crossland, M. D., Sims, M., Galbraith, R. F. & Rubin, G. S. Evaluation of a new quantitative technique to assess the number and extent of preferred retinal loci in macular disease. *Vision Res.* **44**, 1537–1546 (2004).



42. Miller, G. A. & Johnson-Laird, P. N. *Language and Perception@ Language and Perception* (Harvard University Press, Harvard, 2003). <https://doi.org/10.7551/mitpress/10273.003.0008>.
43. Levine, T. Using colour in figures: Let's agree to differ. *Traffic* **10**, 344–347 (2009).
44. Schubert, M. C. & Migliaccio, A. A. Stability of the aVOR to repeat head impulse testing. *Otol. Neurotol.* **37**, 781–786 (2016).
45. Singh, N. K., Govindaswamy, R. & Jagadish, N. Test–retest reliability of video head impulse test in healthy individuals and individuals with dizziness. *J. Am. Acad. Audiol.* **30**, 744–752 (2019).
46. Tobii eye tracking support. How to position the eye tracker and participant in a study. <https://www.tobii.com/learn-and-support/learn/steps-in-an-eye-tracking-study/run/how-to-position-the-participant-and-the-eye-tracker/> (2015).
47. Tobii eye tracking support. What's the difference between Tobii Eye Tracker 4C and Tobii EyeX? [https://help.tobii.com/hc/en-us/articles/360008539058-What-s-the-difference-between-Tobii-Eye-Tracker-4C-and-5-?source=search&auth\\_token=eyJhbGciOiJIUzI1NiJ9.eyJhY2NvdW50X2lkjloxMDM2OTIzLCJ1c2VyX2lkjjozNzk3MzU5MjYxNzcsInRyY2tldF9pZCI6ODgwODIsImNoYX5uZWxfaXo](https://help.tobii.com/hc/en-us/articles/360008539058-What-s-the-difference-between-Tobii-Eye-Tracker-4C-and-5-?source=search&auth_token=eyJhbGciOiJIUzI1NiJ9.eyJhY2NvdW50X2lkjloxMDM2OTIzLCJ1c2VyX2lkjjozNzk3MzU5MjYxNzcsInRyY2tldF9pZCI6ODgwODIsImNoYX5uZWxfaXo) (2020).
48. Kingstone, A. & Klein, R. M. What are human express saccades?. *Percept. Psychophys.* **54**, 260–273 (1993).
49. Holmqvist, K. *et al.* Eye tracking: Empirical foundations for a minimal reporting guideline. *Behav. Res. Methods* **55**, 364–416 (2022).
50. Devillez, H., Guyader, N., Curran, T. & O'Reilly, R. C. The bimodality of saccade duration during the exploration of visual scenes. *Vis. Cogn.* **28**, 484–512 (2020).
51. Abu-Arafah, A., Jordan, H. & Drummond, G. Reporting of method comparison studies: A review of advice, an assessment of current practice, and specific suggestions for future reports. *Br. J. Anaesthesia* **117**, 569–575. <https://doi.org/10.1093/bja/aew320> (2016).
52. Mokkink, L. B. *et al.* COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: A Delphi study. *BMC Med. Res. Methodol* **20**, 1–13 (2020).
53. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163 (2016).
54. Shrout, P. E. Measurement reliability and agreement in psychiatry. *Stat. Methods Med. Res.* **7**, 301–317 (1998).
55. Walter, S. D., Eliasziw, M. & Donner, A. Sample size and optimal designs for reliability studies. *Stat. Med.* **17**, 101–110 (1998).
56. Aickin, M. & Gensler, H. Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods. *Am. J. Public Health* **86**, 726–728 (1996).
57. Fukushima, K. & Kaneko, C. R. S. Vestibular integrators in the oculomotor system. *Neurosci. Res.* **22**, 249–258 (1995).
58. Nakamagoe, K., Iwamoto, Y. & Yoshida, K. Evidence for brainstem structures participating in oculomotor integration. *Science* **1979**(288), 857–859 (2000).
59. King, J. A., McCrea, M. A. & Nelson, L. D. Frequency of primary neck pain in mild traumatic brain injury/concussion patients. *Arch. Phys. Med. Rehabil.* **101**, 89–94 (2020).
60. De Vries, J. *et al.* Cervico-ocular reflex is increased in people with nonspecific neck pain. *Phys. Ther.* **96**, 1190–1195 (2016).
61. Fujiwara, K., Kunita, K. & Furune, N. Effect of vibration stimulation to neck extensor muscles on reaction time in various saccadic eye movements. *Int. J. Neurosci.* **119**, 1925–1940 (2009).
62. Roy, J. E. & Cullen, K. E. Brain stem pursuit pathways: Dissociating visual, vestibular, and proprioceptive inputs during combined eye-head gaze tracking. *J. Neurophysiol.* **90**, 271–290 (2003).
63. Kowler, E., Rubinstein, J. F., Santos, E. M. & Wang, J. Predictive smooth pursuit eye movements. *Annu. Rev. Vis. Sci.* **5**, 223–246 (2019).
64. Lin, A. P. *et al.* Metabolic imaging of mild traumatic brain injury. *Brain Imaging Behav.* **6**, 208–223 (2012).
65. Ettinger, U. *et al.* Reliability of smooth pursuit, fixation, and saccadic eye movements. *Psychophysiology* **40**, 620–628 (2003).
66. Płomecka, M. B., Baranczuk-Turska, Z., Pfeiffer, C. & Langer, N. Aging effects and test–retest reliability of inhibitory control for saccadic eye movements. *Neuro* **7**, 1–16 (2020).
67. Blekher, T. *et al.* Test–retest reliability of saccadic measures in subjects at risk for huntington disease. *Invest Ophthalmol Vis. Sci.* **50**, 5707–5711 (2009).
68. Bargary, G. *et al.* Individual differences in human eye movements: An oculomotor signature?. *Vision Res.* **141**, 157–169 (2017).
69. Clough, M. *et al.* Oculomotor cognitive control abnormalities in australian rules football players with a history of concussion. *J. Neurotrauma* **35**, 730–738 (2018).
70. Kristjánsson, Á., Chen, Y. & Nakayama, K. Less attention is more in the preparation of antisaccades, but not prosaccades. *Nat. Neurosci.* **4**, 1037–1042 (2001).
71. Strickland, T. L., D'Elia, L. F., James, R. & Stein, R. Stroop color-word performance of African Americans. *Clin. Neuropsychol.* **11**, 87–90 (1997).
72. Van Der Elst, W., Van Boxtel, M. P. J., Van Breukelen, G. J. P. & Jolles, J. The stroop color-word test: Influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment* **13**, 62–79. <https://doi.org/10.1177/1073191105283427> (2006).
73. Rodas, J. A. & Greene, C. M. Test–retest reliability of four cognitive tasks in a not so WEIRD sample. *PsyArXiv* **1**, 1–35 (2021).
74. Wang, R. Y., Zhou, J. H., Huang, Y. C. & Yang, Y. R. Reliability of the chinese version of the trail making test and stroop color and word test among older adults. *Int. J. Gerontol.* **12**, 336–339 (2018).
75. Potter, S., Leigh, E., Wade, D. & Fleminger, S. The rivermead post concussion symptoms questionnaire. *J. Neuro.* **253**, 1603–1614 (2006).
76. Menon, D. K., Schwab, K., Wright, D. W. & Maas, A. I. Position statement: Definition of traumatic brain injury. *Arch. Phys. Med. Rehabil.* **91**, 1637–1640. <https://doi.org/10.1016/j.apmr.2010.05.017> (2010).
77. Saatman, K. E. *et al.* Classification of traumatic brain injury for targeted therapies. *J. Neurotrauma* **25**, 719–738 (2008).
78. McGarvie, L. A. *et al.* The video head impulse test (vHIT) of semicircular canal function: Age-dependent normative values of VOR gain in healthy subjects. *Front Neurol.* **6**, 154 (2015).
79. Walker, M. F. & Zee, D. S. Directional abnormalities of vestibular and optokinetic responses in cerebellar disease. in *Annals of the New York Academy of Sciences* vol. 871 205–220 (John Wiley & Sons, Ltd, 1999).

## Author contributions

A.C., P.T. conceived the experiment(s), A.C. conducted the experiment(s), A.C., P.T. analyzed the results. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-63540-8>.

**Correspondence** and requests for materials should be addressed to A.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024