Journal of Science and Medicine in Sport 28 (2025) 377–383

Contents lists available at ScienceDirect

Journal of Science and Medicine in Sport

journal homepage: www.elsevier.com/locate/jsams

Original research

Repeatability, reproducibility and normative ranges for EyeGuide Focus testing system in elite male rugby players

G.W. Fuller [a,*], J. Brown [b], M. Dunlop [c], M. van Rheede [d,e], P. Mathema [f], M. Webb [g], M. Raftery [h], E. Falvey [h]

[a] Centre for Urgent and Emergency Care Research, School of Health and Related Research, University of Sheffield, Regent Court, UK
[b] Institute of Sport and Exercise Medicine, Department of Sport Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa
[c] Edinburgh Rugby, BT Murrayfield Stadium, UK
[d] Sports Rehab Centre, South Africa
[e] UCTRFC (University of Cape Town Rugby Football Club), University of Cape Town, South Africa
[f] Wales Rugby Union, Principality Stadium, UK
[g] Ulster Rugby, Kingspan Stadium, UK
[h] World Rugby, Ireland

ARTICLE INFO

Article history:
Received 19 September 2024
Received in revised form 11 November 2024
Accepted 3 December 2024
Available online 6 December 2024

Keywords:
Rugby
Repeatability
Reliability
Reproducibility
Smooth pursuits
Concussion

ABSTRACT

Objectives: The EyeGuide Focus system is a simple, portable, test of visual tracking with potential use for concussion screening. This study investigated the repeatability, reproducibility, distribution, and modifiers of EyeGuide Focus measurements in healthy elite Rugby players.
Design: Cross sectional repeated measures study and controlled pre-test post-test sub-study.
Methods: EyeGuide Focus testing was performed in a medical room at rest. Test-retest repeatability (within-subject standard deviation (Sw), coefficient of variation (CV), repeatability coefficient (RC)) and reliability intra-class correlation coefficient (ICC A,1) of 3 test results were evaluated. The distribution of best score across replicates was then examined using summary statistics, and the influence of subject characteristics investigated. A controlled pre-test post-test sub-study examined the effect of exercise on best EyeGuide Focus score using an analysis of covariance (ANCOVA).
Results: A total of 769 elite male Rugby players underwent EyeGuide Focus testing. Repeated test scores demonstrated positively skewed distributions. Test-retest repeatability (Sw 1.46, CV 46.0 %, RC 2.85, natural log transformed data) and reliability (ICC 0.41, natural log transformed data) were low. The distribution of best EyeGuide Focus score was unaffected by previous concussion, eye conditions, or age. No learning or exercise effects were apparent in the controlled pre-test post-test sub-study (ANCOVA, $n = 89$, $p = 0.69$).
Conclusions: EyeGuide Focus test-retest repeatability and reliability were low and could limit diagnostic accuracy. Best test score achieved across repeated measurements is a possible metric for operationalisation and future research is required to determine if this differs in patients with concussion compared to normal subjects.

Practical implications

- Smooth pursuit eye movements may be affected by head impacts and a visual tracking task could be a useful test paradigm for assessment of sport-related concussion.
- The EyeGuide Focus system is a simple, portable, test of visual tracking.
- Overall EyeGuide Focus test-retest repeatability and reliability were low, but comparable with established concussion screening tests.
- Best score achieved across repeated measurements appears to be a possible metric for future clinical operationalisation.
- Future research will determine if the distribution of best score in patients with concussion differ from those in normal subjects.

1. Introduction

Concussion is a common and high-profile injury in contact and collision sports. Short term sequalae include somatic, cognitive and neurological symptoms or signs, increased risk of injury, and reduced athletic performance.[1] Early detection of suspected concussion and

* Corresponding author.
E-mail addresses: g.fuller@sheffield.ac.uk (G.W. Fuller), michael.dunlop@edinburghrugby.org (M. Dunlop), pmathema@wru.wales (P. Mathema), webbm@doctors.org.uk (M. Webb), Martin.Raftery@worldrugby.org (M. Raftery), Eanna.Falvey@worldrugby.org (E. Falvey).

https://doi.org/10.1016/j.jsams.2024.12.002
1440-2440/

removal of the affected player from the field of play will likely mitigate these potential adverse effects and facilitate further evaluation, management, and safer return-to-play of the injured player.[2]

Smooth pursuit eye motion, consisting of saccadic and tracking ocular movements, is designed to keep a moving stimulus on the fovea, supporting perceptual stability of a moving object of interest.[3] Visual tracking metrics have been associated with cognitive functioning and integrity of frontal white matter tracts that are vulnerable to a concussive impact.[4] Therefore, a visual tracking task could be a useful test paradigm for side-line screening, diagnosis, or monitoring recovery of sport-related concussion.[4,5]

The EyeGuide Focus system is a simple, portable, test of visual tracking.[6] This study aimed to determine the technical performance of the EyeGuide system prior to future evaluation of diagnostic accuracy. Specific objectives were to determine the repeatability and reproducibility of EyeGuide measurements in healthy elite Rugby players, investigate the effects of subject characteristics on performance, and establish normative test ranges.

2. Methods

Players from the 2021/2022 United Rugby Championship (URC) elite male Rugby Union competition were studied. A cross sectional repeated measures study was conducted on all first team squad Rugby players registered with each of the 16 URC teams to examine test-retest repeatability, reliability, investigate the influence of subject characteristics and establish normative ranges for EyeGuide Focus results. An ancillary controlled pre-test post-test reproducibility sub-study to investigate the possibility of exercise effects was also conducted in a separate convenience sample, comprising players from 12 URC teams, and a single South African University first Rugby team.

The index test under investigation was the EyeGuide Focus system (EyeGuide, Lubbock, Texas, United States, Fig. 1).[6] EyeGuide Focus measures "smooth pursuit" or Dynamic Visio-Motor Synchronization (DVS) as characterised by Maruta and colleagues.[4] Briefly, subjects look at a screen and focus on a small white target stimulus, set on a black background. The target moves in a horizontal 'lazy eight' pattern, starting at the centre of the display and moving clockwise at 0.4 Hz in a circular trajectory on the right side of the screen. When the stimulus returns to the centre of the display, the path changes to a counter-clockwise circle on the left side of the screen. The test ends when the stimulus returns to the centre of the display. A built-in digital camera tracks the pupil centre coordinates during the task, and these are compared to the movements

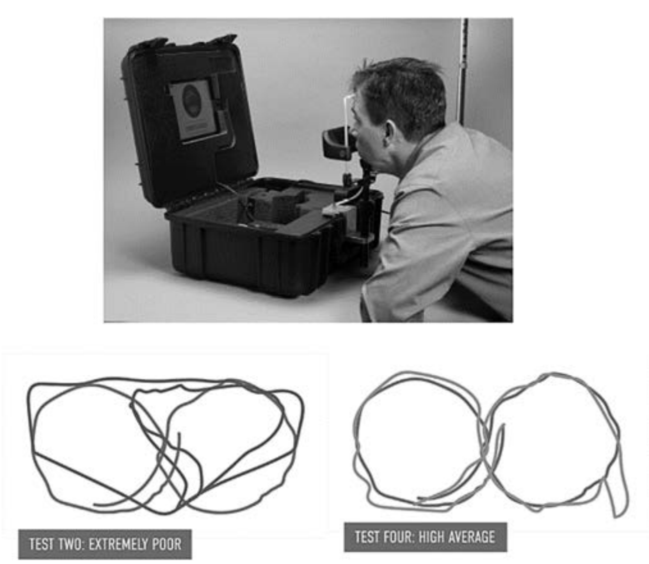

Fig. 1. EyeGuide Focus equipment (upper panel) and records of user eye movements during 10 s smooth pursuit test (lower panel).

Table 1
Subject characteristics.

Subject characteristic	Summary statistic (n = 723)	
	Point estimate	Spread
Age (years)	Median 24.9	IQR 22.3–28.8
Years of education (years)	Median 14	IQR 12–16
Previous concussion (n, %)	355 (49.1 %)	
Number of concussions (n, %)	0: 368 (50.9 %)	
	1: 112 (15.5 %)	
	2: 98 (13.6 %)	
	3: 61 (8.4 %)	
	4: 27 (3.7 %)	
	5: 22 (3.3 %)	
	>5: 35 (4.8 %)	
Eye condition (n, %)		
-Refractive error	48 (6.6 %)	
-Astigmatism	20 (2.8 %)	
-Detached retina	2 (0.2 %)	
Replicate 1	Median 22,263	IQR 15,747–34,817
Replicate 2	Median 20,725	IQR 15,044–34,282
Replicate 3	Median 21,298	IQR 14,342–34,005

IQR: inter-quartile range.

required to remain focussed on the moving circle (Fig. 1). The test lasts exactly 10 s, with pupil centre coordinates measured 60 times per second. A final summary result is provided, calculated as the sum of the distances between the pupil centre coordinates and the actual on-screen stimulus coordinates for the duration of the test except the first and last seconds. Thus, a score of 0 indicates flawless performance through the operational duration of the test, whilst higher scores indicate worse performance. Should the camera be unable to track the movement of the pupils due to a reflection or any other reason, the test aborts due to a 'pupil lock' failure. Operationally EyeGuide Focus routinely imposes an upper limit for an eligible test score of 38,842, corresponding to two standard deviations above the mean value of 22,651 from a large sample of uninjured community controls (Shane Keating, EyeGuide, personal communication, 2nd May 2024).

To assess repeatability, reliability, and investigate the influence of subject characteristics, and establish normative ranges all squad players participating in the URC competition underwent EyeGuide testing prior to commencement of the 2021/2022 competition season. Standardised testing was performed at rest in a medical room prior to a training session, followed EyeGuide Focus recommendations, and was administered by a separate trained assessor in each separate team. Subjects could undergo up to 6 attempts ('trials') to achieve 3 successful EyeGuide Focus measurements ('replicates'). Trials were performed consecutively with no intervening break. Successful replicates were defined as a test without a 'pupil lock failure' where any pupil tracking measurement was recorded. From these replicates, eligible test scores of <38,842 were included in subsequent analyses. The number of trials and replicates was chosen to reflect the limited time available for sideline concussion screening. Analysis of replicate measurements and eligible test scores was performed separately to evaluate both the full range of potential measurements and the operational performance of EyeGuide Focus. Testing was performed under each player's usual training conditions i.e. contact lens/glasses used/not used as normal.

Relevant demographic and clinical information was recorded using a standardised electronic recording form by team assessors prior to testing. EyeGuide Focus performance data were collected automatically at the time of testing by the secure, electronic EyeGuide Focus system. As the EyeGuide system does not keep a record of pupil lock fails, all testers were asked to document the sequence and number of pupil lock fails experienced by subjects in a separate electronic recording sheet.

The main analysis proceeded in 6 stages. Firstly, derivation of the study sample, describing 'trials' (attempts at measurement), 'replicates' (any recordable EyeGuide measurement) and eligible 'test scores' (measurements below the operational limit of 38,842), was enumerated. Demographic characteristics of the sample were then described.

Secondly, the distribution of EyeGuide Focus measurements across repeated replicates was summarised visually, with medians and interquartile ranges used to describe distributions. Friedman's One-Way Repeated Measure Analysis of Variance by Ranks was then used to test for any significant changes in median measurements across replicates. Nemenyi's all-pairs comparisons tests were planned for post-hoc

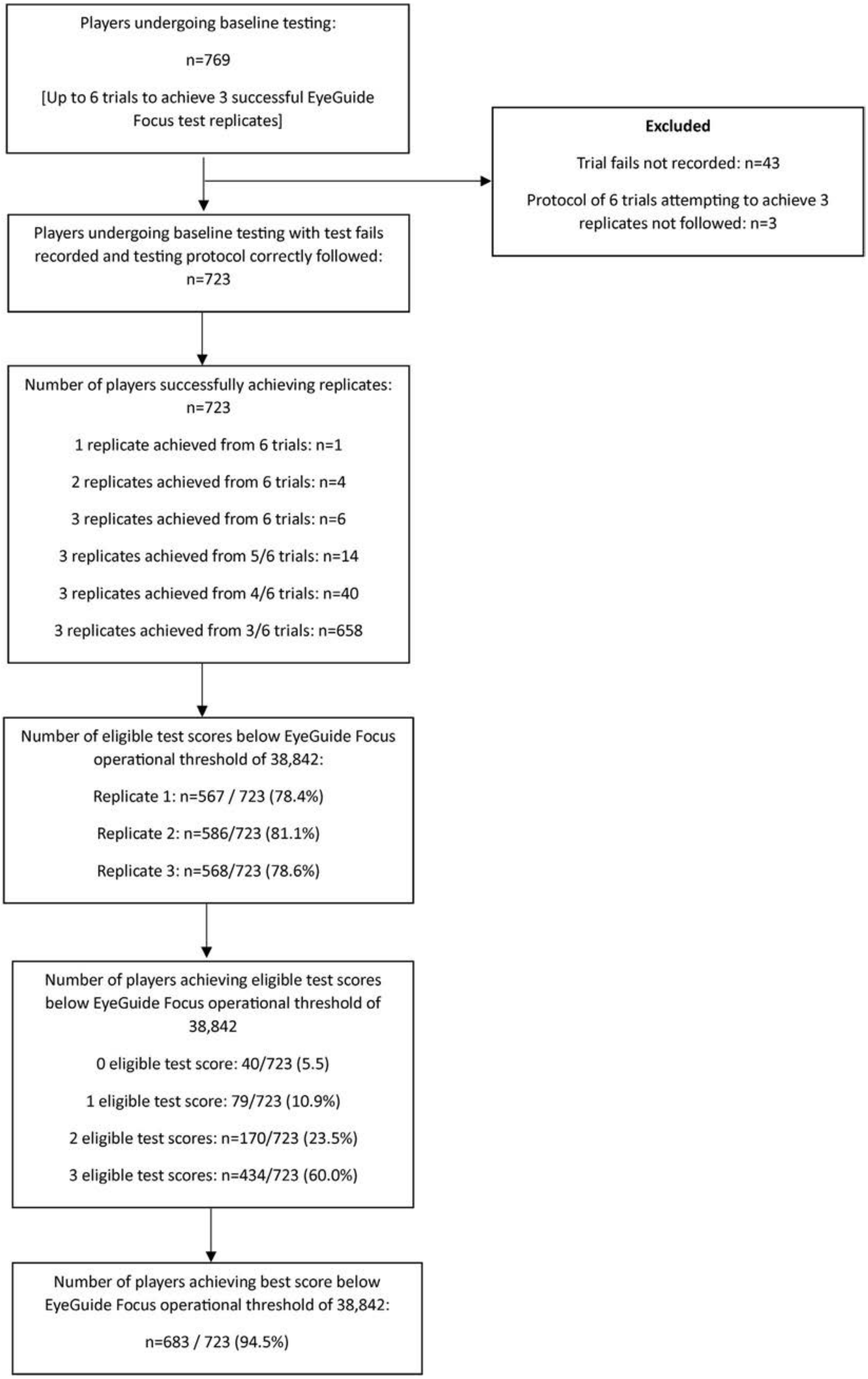

Fig. 2. Flow chart demonstrating derivation of the study population.

comparisons if required. Replicate measurements for each subject were also plotted in order of decreasing within-subject variability.

Thirdly, distributions of repeated test scores (i.e., measurements below 38,842) were examined. Test-retest repeatability, and reliability of repeated test scores were then evaluated. Test scores were transformed for these analyses using natural logarithms to achieve normality and provide independence between the subject log standard deviation subject log mean (see supplementary materials for details).[7–9] Within-subject standard deviation (also known as typical error) was then estimated from a one-way repeated analysis of variance (ANOVA). Within subject coefficient of variation (CV%) was used to provide a standardised measure of the average within-subject variability of repeated tests, allowing comparison with other tests and subject groups.[9–12] The repeatability coefficient was then used to indicate the maximum difference likely to occur between two successive test measurements within the same subject.[9–11]

Test-retest reliability of natural logarithm transformed score data was evaluated through calculation of an intraclass correlation coefficient (ICC) based on a single rater, absolute-agreement, 2-way mixed-effects model (ICC A,1).[13,14] The guidelines given by Koo were used to assess the ICC coefficients: 0.90–0.99: high reliability, 0.80–0.89: good reliability, 0.70–0.79: fair reliability, and ≤0.69: poor reliability.[13]

Fourthly, a summary statistic approach was taken using the best test score across replicates as a single metric for operationalisation of EyeGuide Focus testing. The distribution of untransformed best test score was characterised graphically and using descriptive statistics. Normative ranges were then determined with cutoffs selected based on distribution percentiles consistent with previous sport medicine normative values.[15] As the distribution of best score was again positively skewed, these data were similarly transformed using the natural logarithm. Differences in the distribution of best score across different teams were then assessed used a one-way ANOVA and Tukey post-hoc testing.

Fifthly, differences in the distribution of best EyeGuide Focus test score between healthy subjects and those with previous concussion or refractive eye problems were examined using Mann–Whitney *U* tests. 95 % confidence interval for difference in median score was calculated using 1000 bootstrap replicates. The association between best score and age and years of education was investigated using scatter plots and Spearman's correlation coefficient.

Sixthly, in the controlled pre-test, post-test, ancillary sub-study players were allocated in a quasi-randomised fashion, according to availability of the assessor, to one of two study arms: a) intervention arm: rest-exercise; and b) control arm: rest-rest. The aim of this study was to determine if there were any effects of exercise on EyeGuide Focus testing, with a null hypothesis that measurements would not significantly differ after exercise compared to testing in the resting state. In both arms participants underwent testing in a gymnasium in a distraction-free environment at rest as described above. The intervention arm then completed a 10-minute standardised exercise protocol using an exercise bike. Participants were required to exercise to 70–80 % of their age predicted maximum heart rate and reach a target Rate of Perceived Exertion Scale of 7 or greater.[16,17] EyeGuide Focus testing (post-exercise state) subsequently occurred as described previously within 2 min of completing the exercise protocol. Instead of the exercise protocol, control arm participants waited 8 min before repeating the standardised EyeGuide Focus testing at rest. An analysis of covariance (ANCOVA), using natural logarithm transformed data, was then used to determine the effect of exercise on the best EyeGuide test score from 3 trials after controlling for basal best EyeGuide score of subjects.

All other assumptions for statistical tests were checked and verified. Statistical significance was defined as a p value <0.05. Family-wise type 1 error rate was maintained with a Bonferroni correction for multiplicity if multiple statistical comparisons were made. Statistical analyses were carried out using R 4.1.2[33] using the R Studio interface and Stata v18 (StataCorp. 2023. Stata Statistical Software: Release 18. College Station, TX: StataCorp LLC).

As the study population consisted of a census sample of all available players in the URC competition, the final sample size was fixed. For the pre-test, post-test, controlled study sample size calculation was performed in G*Power 3.1.9.6 (Universität Kiel, Germany).[18] Assuming power of 0.8, level of significance of 0.05, and effect size of 0.3, an ANCOVA analysis would require 90 subjects. A study protocol with an a priori investigation plan was developed prior to analysis. The investigation plan received ethical approval from an independent World Rugby Ethics Board. Study participants provided written informed consent for participation and use of anonymised data.

3. Results

From a total of 791 rostered players, 769 elite male Ruby players underwent EyeGuide Focus testing (97.2 %). Of these, 46 participants were excluded as the testing protocol was not followed correctly. Most of the remaining athletes successfully completed 3 replicates without any pupil lock fails (658/723, 91.0 %). From these 94.5 % (683/723) of players achieved more than one test score below the EyeGuide Focus operational threshold; with 68.9 % (498/723) of players achieving 3 eligible test scores. Subject characteristics are summarised in Table 1. Derivation of the study sample is presented in Fig. 2.

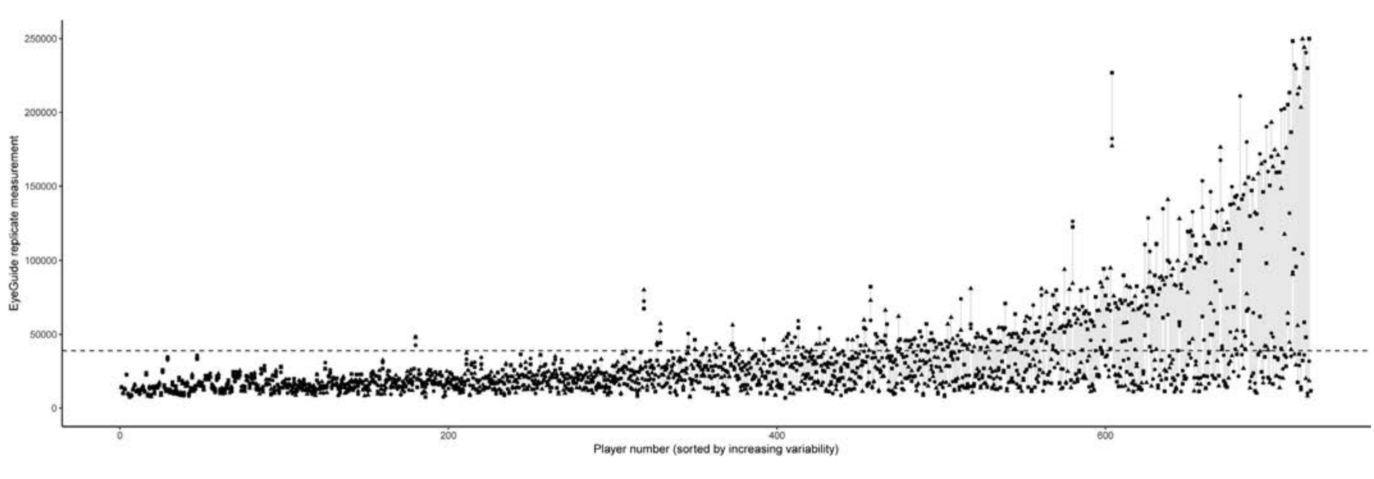

Fig. 3. Test-retest repeatability of EyeGuide Focus measurements. Scatter plot of EyeGuide Focus replicates for each subject ordered according to within-subject variance. Upper/middle/lower panels = Third of subjects with lowest/medium/highest repeatability respectively. Trial 1 = squares; trial 2 = circles; trial 3 = triangles. Dotted line represents median best score. Approximately one third of subjects have low intra-subject variability on repeated testing (top panel), with variability increasing as mean performance worsens. However, even in subjects with low repeatability, a relatively good best score is usually evident.

Table 2
Normative values for best EyeGuide Focus test scores.

	Centiles[a]				
	≤2 %	3–10 %	10–89 %	90–97 %	≥98 %
Indicative performance[a]	Very good	Good	Broadly normal	Poor	Very poor
Cut point	≤8398	8399–10,137	10,138–27,452	27,453–35,997	≥35,998

[a] Centiles and indicative performance consistent with previous sport medicine normative ranges.

EyeGuide Focus replicates demonstrated extremely positively skewed distributions, with median scores of 22,263, 20,725, and 21,298 across the 3 repeated measurements (Table 1, see supplementary materials for further details). Friedman's test showed that there was no significant change in the median of replicate measurements ($p = 0.05$). Within player variability on repeated testing appeared to vary across distinct subgroups of subjects, with approximately one third demonstrating low, moderate and high variability on repeat testing respectively (Fig. 3).

Untransformed test scores also demonstrated positively skewed distributions, with medians of 19,392, 18,809, and 18,033, across repeated measurements (Table 1, supplementary materials). Friedman's test indicated a significant difference in score ranks ($p < 0.01$) with the first test score being significantly higher than subsequent measurements (adjusted $p = 0.01–0.02$).

Overall, within-subject coefficient of variation for natural logarithm transformed test scores was 46.0 % (natural log transformed data, 95 % CI 43.2–48.9 %). Within-subject geometric standard deviation was 1.46 (natural log transformed data, 95 % CI 1.43–1.49), resulting in a repeatability coefficient of 2.85 (natural log transformed data, 95 % CI 2.70–3.01). Test-retest reliability was low with an ICC of 0.41 (natural log transformed data, 95 % CI 0.35–0.47).

The distribution of untransformed best score across eligible test scores similarly followed a positively skewed distribution, with a median score of 14,931 (range 7646–38,174, IQR 11,994–19,418; see supplementary materials). Normative ranges are shown in Table 2. A statistically significant difference in mean best score was apparent across teams (log transformed data, $p < 0.01$). Post-hoc testing revealed a single team with statistically significant worse performance than 3 other teams (mean of log transformed best test score = 9.88 versus 9.54, 9.60, 9.60, $p \leq 0.01$; see supplementary materials for further details).

There were no differences in the distribution of best Eye Guide Focus test scores between normal subjects and those with previous concussions or requiring refractive correction ($p = 0.79$ and 0.52 respectively, see supplementary materials for further details). No association was evident between age, or years of education, and best score ($p = 0.13$, 0.49, see supplementary materials for further details).

In total, 89 players participated in the pre-test, post-test, controlled sub-study (URC n = 72, University n = 17 players). Of these 38 were allocated to the control rest-rest arm and 51 allocated to the intervention rest-exercise arm. There were 15 players who were unable to achieve an eligible test score in the pre- and/or post-testing condition (3 rest arm, 12 exercise arm). Pre-test log-transformed mean best test scores were 9.91 and 9.86 in the exercise and control groups respectively. After adjustment for basal EyeGuide Focus best test score, there were no differences in post-exercise best test scores between each arm (log-transformed data, $p = 0.69$). Adjusted post-test mean best scores were 9.88 (log-transformed data, 95 % CI 9.7–10.0) and 9.85 (95 % CI 9.8–10.0) in the exercise and control groups respectively (see supplementary materials for further details).

4. Discussion

In a large sample of elite male Rugby Union players repeated EyeGuide Focus smooth pursuit measurements demonstrated large within-player variability, with more divergent measurements as subject performance worsened. Following truncation of measurements according to operational thresholds and natural log transformation, overall test-retest repeatability (within-subject SD geometric 1.46, CV 46.0 %, repeatability coefficient 2.85) and reliability (ICC A,1 = 0.41) were low. There were no learning or exercise effects apparent, and the distribution of best test score did not vary according to baseline characteristics of previous concussion, eye conditions, educational level, or age. However, there were differences in the distribution of best test score across different teams.

EyeGuide Focus measurement error was proportional to the mean, denoting a multiplicative structure for the errors. The within-subject geometric standard deviation of 1.46, indicates the difference between a subject's given measurement and their 'true' average value over all possible measurements would be 2.86 times higher, or lower, for 95 % of observations. The repeatability coefficient of 2.85, suggests that repeated within-subject measurements will be <2.85 times apart, for 95 % of pairs of observations. The CV of 46.0 % denotes the within-subject standard deviation approximately equals half the mean. The ICC A,1 of 0.41 implies that measurement errors are large in comparison to the true differences between subjects. This variability in repeated EyeGuide Focus test measurements could reduce its diagnostic accuracy and limit clinical utility.

Differences in repeated measurements made on the same subject under identical conditions may be due to natural variation in the subject, or variability in the measurement process.[11] Marked short-term fluctuation in visual tracking ability seems biological implausible, suggesting that an aspect of the testing process is responsible for the observed lack of repeatability. Interestingly, performance appeared to vary across distinct participant subgroups. Whilst some subjects recorded repeated measurements within a relatively narrow range, approximately one third demonstrated increasingly low repeatability as mean performance worsened. Possible reasons for such measurement error could include idiosyncratic difficulty in 'locking on' to the stimulus, head movement during testing, distractibility, or unknown subject characteristics. Overall, 5.5 % of players were unable to achieve an eligible test score below the operational threshold despite repeated trials. Consequently, it is possible that a minority of subjects may not be amenable to sideline assessment with the EyeGuide Focus system.

A relatively narrow between-subject spread was evident in the summary statistic of best score. Even in subjects with low repeatability, a relatively good best score was usually achieved. Operationalisation, as the best score in 3 successful replicates could therefore be a possible testing paradigm requiring further study. Such an approach has been used in other clinical settings, for example peak expiratory flow rate in asthma.[19] However, it should be noted that further research is required to confirm within-player reliability for best score, and stability of best score across different testing occasions. Alternatively, due to the wide within-subject variation on repeated testing, interpretation of Eye Guide Focus scores may be facilitated by log transformation, analogous to the Richter scale or pH measurement.[20]

The repeatability and reproducibility of an overall quantitative summary of visual tracking provided by EyeGuide Focus were studied. However, individual traces of eye tracking are also recorded and are available for interpretation. Changes to qualitative aspects of visual tracking, such as atypical or vertical/horizontal saccades, may occur following traumatic brain injury and could provide more detailed diagnostic information in sport-related concussion. Further study is required to investigate the utility of EyeGuide pupil traces compared to physical examination of tracking and saccades. It is also possible that the fraction of test fails differs following concussion, compared to normal subjects, and this could represent a further outcome of interest.

Best test score remained stable across baseline characteristics and in different testing conditions, with no exercise or short-term learning effects apparent. However, there was a significant difference in the distribution of best test score between a single team and three others. Given the objective, and automatic nature of the EyeGuide Focus system, low between-team variability might be expected. Detailed information on testing conditions (e.g., time of day, ambient light, noise levels, etc.) was not recorded and the reason for this difference is unclear. The clinical significance of the relatively small observed difference is also uncertain. Formal inter-observer studies, and investigation of stability of best score in the longer term would further increase confidence in reproducibility.

The aim of a phase 0 diagnostic study is to establish the technical properties of an index test in terms of its validity (i.e., the extent to which a test measures what it is intended to measure) and precision (i.e., the consistency of any variation of the test measurements with the true variations).[21] Content, construct, and criterion validity of smooth pursuit tests have been previously established.[22,23] Characterisation of precision in this study allows progression to phase 1 diagnostic research investigating if EyeGuide Focus results differ between concussed and healthy subjects. However, given the apparent low repeatability and reliability, the ability of EyeGuide Focus to accurately discriminate could be questioned. If proof of concept is established in case-control studies, future single gate diagnostic cohort and impact studies can examine accuracy and effectiveness in side-line concussion screening and diagnosis.[21]

The large study sample, comprising players of different ages, and from a wide range of countries, educational levels, and playing positions should ensure excellent external validity within male professional Rugby Union. The testing protocol, using a trained assessor in a medical room setting, should allow extrapolation of findings to current in-match concussion screening procedures used in professional Rugby Union. Baseline best EyeGuide Focus score could therefore be investigated as an internal comparator in the Head Injury Assessment process, rather than using normative cut-points. Generalisability to other sports, amateur Rugby Union, and female or younger subjects is less certain. The controlled pre-test post-test sub-study used a bicycle protocol to represent exercise effects. Although plausible levels of exertion were achieved (Perceived Exertion Scale of 7) it is possible that results could differ in the context of typical Rugby practice or play.

There has been limited research into the reproducibility of EyeGuide Focus measurements. Pearce and colleagues investigated the test-retest reliability of 3 repeated EyeGuide Focus measurements, from up to 10 attempted trials, in 75 healthy adolescents and adults, reporting a favourable ICC of 0.79 (95 % CI 0.70–0.86).[24] This discordant result may be explained by the restricted study population of health subjects, different testing paradigm (up to 10 trials), and lack of data transformation. In contrast, Walshe and colleagues observed a low ICC of 0.27 (95 % CI 0.04–0.51) from unselected field athletes which are more consistent with the current study.[25] In other EyeGuide Focus studies, Kelly investigated 849 junior athletes, reporting a median score of 28,373 (IQR 13,263) from a single trial; broadly comparable to the results of individual trials observed in the current study,[6] although technical and methodical differences in data collection may limit direct comparison. Possible utility of visual tracking in side-line assessment of head impacts was also provided by Kelly who reported a statistically significant difference in the distribution of baseline and follow-up scores in 42 subjects diagnosed with concussion ($p < 0.01$).[6] These findings are supported by Still and colleagues who studied mixed martial art fighters before and after bouts, demonstrating a statistically significant change from a pre-fight baseline test mean of 17,426 to a post-fight mean of 37,694 ($p < 0.01$).[26] Several other video-oculography devices to test visual tracking have been described, but comparison to the current findings is difficult due to different measurement scales.[27]

Benchmarking EyeGuide Focus performance against other concussion screening tests is limited by differing assessment periods and methodology but suggests similarly low test-retest repeatability and reliability. Broglio and colleagues reported 'less than optimal reliability for most common and emerging concussion assessment tools' in a large cohort of college athletes retested 1 year apart.[28] More recently, the Sports Concussion Assessment Tool 5 (SCAT5) was demonstrated to have 'marginal or low' 2-week retest reliability across constituent subtests.[29,30] For example, the observed ICC of 0.49 for the Standardised Assessment of Concussion is lower than that reported herein. Other oculomotor tests have also demonstrated large repeatability coefficients, for example the King-Devick test ($\pm$8.76 s), albeit in adolescent subjects.[31]

This study has several strengths. Standardised testing was performed by trained medical practitioners and the electronic data collection system allowed immediate data collection without missing data. Conversely there are potential limitations. A small minority of players were excluded from the study. Although the missingness mechanism appears to be missing completely at random (e.g. COVID-19 isolation) selection bias cannot be completely excluded. There were several subject characteristics that were not measured, but which could affect EyeGuide Focus scores, such as prescription drug use, lack of sleep, and skin tone. Other testing conditions, including distractions, and reproducibility of scores over the longer term, were also not studied. Finally, the pre-test, post-test, controlled sub-study was unblinded, and did not formally randomise subjects to each study arm. Although unlikely given the homogenous study population and objective testing procedures, systematic error from selection and information biases is possible.[32]

This study reports on the precision of EyeGuide Focus, a low-cost, ocular-based test, for assessment of visual tracking. Overall test-retest repeatability and reliability of test scores were low. Although not dissimilar to established concussion screening tests, this lack of precision could limit diagnostic accuracy. Best test score achieved across repeated measurements is a possible metric for operationalisation and future research will determine if this differs in patients with concussion compared to normal subjects.

CRediT authorship contribution statement

GF, EF, and MR conceived the study. GF designed the study. JB centrally managed the study. MD locally managed the study. MD, MW and JB monitored and coordinated data collection. MR and MP performed the controlled pre-test post-test sub-study. GF performed the analyses. All authors made substantial contributions to the design, data processing, and interpretation. GF drafted the article and all other authors revised it critically for important intellectual content. EF is the guarantor. All authors had full access to all the data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

Confirmation of ethical compliance

The authors declare that the procedures were followed according to the regulations established by the Helsinki Declaration of the World Medical Association. The investigation plan received ethical approval from an independent World Rugby Ethics Board. Study participants provided written informed consent for participation and use of anonymised data.

Funding information

The study was supported by World Rugby. EyeGuide (EyeGuide, Lubbock, Texas, United States) supplied testing units and provided training, but otherwise was not involved in the study.

Declaration of interest statement

JB's Postdoctoral Fellowship at Stellenbosch University is partially funded by World Rugby. EF and MR are employed by World Rugby. GF has received travel expenses to attend scientific meetings from World Rugby. The authors declare no other conflicts of interest.

Acknowledgements

We would like to acknowledge and sincerely thank the team doctors and EyeGuide testers for their help in facilitating collection of EyeGuide Focus data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsams.2024.12.002.

References

1. Harmon KG, Drezner JA, Gammons M et al. American Medical Society for Sports Medicine position statement: concussion in sport. *Br J Sports Med* 2013;47(1):15-26. doi:10.1136/bjsports-2012-091941.
2. McCrory P, Meeuwisse WH, Aubry M et al. Consensus statement on concussion in sport: the 4th International Conference on Concussion in Sport held in Zurich, November 2012. *Br J Sports Med* 2013;47(5):250-258. doi:10.1136/bjsports-2013-092313. [published Online First: 2013/03/13].
3. Behling S, Lisberger SG. Different mechanisms for modulation of the initiation and steady-state of smooth pursuit eye movements. *J Neurophysiol* 2020;123(3):1265-1276. doi:10.1152/jn.00710.2019. [published Online First: 2020/02/20].
4. Maruta J, Suh M, Niogi SN et al. Visual tracking synchronization as a metric for concussion screening. *J Head Trauma Rehabil* 2010;25(4):293-305. doi:10.1097/HTR.0b013e3181e67936. [published Online First: 2010/07/09].
5. Bin Zahid A, Hubbard ME, Lockyer J et al. Eye tracking as a biomarker for concussion in children. *Clin J Sport Med* 2020;30(5):433-443. doi:10.1097/jsm.0000000000000639. [published Online First: 2018/08/11].
6. Kelly M. Technical report of the use of a novel eye tracking system to measure impairment associated with mild traumatic brain injury. *Cureus* 2017;9(5). doi:10.7759/cureus.1251. [e1251-e51].
7. Bland JM, Altman DG. Statistics notes: logarithms. *BMJ* 1996;312(7032):700. doi:10.1136/bmj.312.7032.700.
8. Bland JM, Altman DG. Statistics notes: transforming data. *BMJ* 1996;312(7033):770. doi:10.1136/bmj.312.7033.770.
9. Bland JM, Altman DG. Statistics notes: measurement error proportional to the mean. *BMJ* 1996;313(7049):106. doi:10.1136/bmj.313.7049.106.
10. Bland JM, Altman DG. Measurement error proportional to the mean. *BMJ (Clin Res Ed)* 1996;313(7049):106. doi:10.1136/bmj.313.7049.106.
11. Bland JM, Altman DG. Measurement error. *BMJ (Clin Res Ed)* 1996;312(7047):1654. doi:10.1136/bmj.312.7047.1654.
12. Canchola JA. Correct use of percent coefficient of variation (%CV) formula for log-transformed data. *MOJ Proteom Bioinform* 2017;6(3). doi:10.15406/mojpb.2017.06.00200.
13. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012. [published Online First: 2016/03/31].
14. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;1(1):30-46. doi:10.1037/1082-989x.1.1.30.
15. Fuller GW, Raftery M. Sport concussion assessment tool: fifth edition normative reference values for professional rugby union players. *Clin J Sport Med* 2020;30(5):e150-e153. doi:10.1097/jsm.0000000000000713. [published Online First: 2018/12/28].
16. Arena R, Myers J, Kaminsky LA. Revisiting age-predicted maximal heart rate: can it be used as a valid measure of effort? *Am Heart J* 2016;173:49-56. doi:10.1016/j.ahj.2015.12.006. [published Online First: 2016/02/28].
17. Borg GA. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc* 1982;14(5):377-381. [published Online First: 1982/01/01].
18. Faul F, Erdfelder E, Lang AG et al. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007;39(2):175-191. doi:10.3758/bf03193146 [published Online First: 2007/08/19].
19. Frischer T, Meinert R, Urbanek R et al. Variability of peak expiratory flow rate in children: short and long term reproducibility. *Thorax* 1995;50(1):35-39. doi:10.1136/thx.50.1.35. [published Online First: 1995/01/01].
20. Boore DM. The Richter scale: its development and use for determining earthquake source parameters. *Tectonophysics* 1989;166(1–3):1-14.
21. Colli A, Fraquelli M, Casazza G et al. The architecture of diagnostic research: from bench to bedside—research guidelines using liver stiffness as an example. *Hepatology* 2014;60(1):408-418. doi:10.1002/hep.26948. [published Online First: 2013/11/28].
22. Mani R, Asper L, Khuu SK. Deficits in saccades and smooth-pursuit eye movements in adults with traumatic brain injury: a systematic review and meta-analysis. *Brain Inj* 2018;32(11):1315-1336. doi:10.1080/02699052.2018.1483030. [published Online First: 2018/06/19].
23. Fukushima K, Fukushima J, Warabi T et al. Cognitive processes involved in smooth pursuit eye movements: behavioral evidence, neural substrate and clinical correlation. *Front Syst Neurosci* 2013:7. doi:10.3389/fnsys.2013.00004.
24. Pearce AJ, Daly E, Ryan L et al. Reliability of a smooth pursuit eye-tracking system (EyeGuide Focus) in healthy adolescents and adults. *J Funct Morphol Kinesiol* 2023;8(2):83.
25. Walshe A, Daly E, Pearce AJ et al. In-season test-retest reliability of visual smooth-pursuit (EyeGuide Focus) baseline assessment in female and male field-sport athletes. *J Funct Morphol Kinesiol* 2024;9(1). doi:10.3390/jfmk9010046. [published Online First: 20240304].
26. Still BBB, Kelly M. Evaluating the efficacy of smooth pursuit visu https://uk.eyeguide.tech/casestudy/case-study-title-3/ 2019. Accessed 21 April 2022.
27. Stuart S, Parrington L, Martini D et al. The measurement of eye movements in mild traumatic brain injury: a structured review of an emerging area. *Front Sports Act Living* 2020;2:5. doi:10.3389/fspor.2020.00005.
28. Broglio SP, Katz BP, Zhao S et al. Test-retest reliability and interpretation of common concussion assessment tools: findings from the NCAA-DoD CARE consortium. *Sports Med (Auckland, NZ)* 2018;48(5):1255-1268. doi:10.1007/s40279-017-0813-0.
29. Bruce JM, Meeuwisse W, Hutchison MG et al. Determining Sport Concussion Assessment Tool fifth Edition (SCAT5) reliable change in male professional hockey players. *Br J Sports Med* 2022. doi:10.1136/bjsports-2021-104851. [published Online First: 20220325].
30. Hänninen T, Parkkari J, Howell DR et al. Reliability of the Sport Concussion Assessment Tool 5 baseline testing: a 2-week test-retest study. *J Sci Med Sport* 2021;24(2):129-134. doi:10.1016/j.jsams.2020.07.014. [published Online First: 20200805].
31. Oberlander TJ, Olson BL, Weidauer L. Test-retest reliability of the King-Devick test in an adolescent population. *J Athl Train* 2017;52(5):439-445. doi:10.4085/1062-6050-52.2.12. [published Online First: 20170331].
32. Sterne JAC, Savović J, Page MJ et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898. doi:10.1136/bmj.l4898.
33. R Core Team. *R: A language and environment for statistical computing,* Vienna, Austria, R Foundation for Statistical Computing, 2021https://www.R-project.org/.