

Article

Advancing Cognitive–Motor Assessment: Reliability and Validity of Virtual Reality-Based Testing in Elite Athletes

Cathy Craig ^{1,*}, Erin Noble ², Mario A. Parra ² and Madeleine A. Grealley ²¹ School of Psychology, Ulster University, Coleraine BT52 1SA, UK² Department of Psychological Sciences and Health, University of Strathclyde, Glasgow G1 1QE, UK;

erin.noble@strath.ac.uk (E.N.); mario.parra-rodriguez@strath.ac.uk (M.A.P.); m.grealley@strath.ac.uk (M.A.G.)

* Correspondence: c.craig1@ulster.ac.uk

Abstract

Emerging virtual reality (VR) technologies provide objective and immersive methods for assessing cognitive–motor function, particularly in elite sport. This study evaluated the reliability and validity of VR-based cognitive–motor assessments in a large sample of elite male athletes ($n = 829$). Ten cognitive–motor tests, delivered via Oculus Quest 2 headsets, were used, covering four domains: Balance and Gait (BG), Decision-Making (DM), Manual Dexterity (MD), and Memory (ME). A Confirmatory Factor Analysis (CFA) was conducted to establish a four-factor model and generate data-driven weights for domain-specific composite scores. The results demonstrated that the composite scores for BG, MD, ME, and a Global Cognitive–Motor (CM) score were all normally distributed. However, the DM score significantly deviated from normality, exhibiting a pronounced ceiling effect. Test–retest reliability was high across all cognitive–motor domains. In summary, VR assessments offer ecologically valid and precise measurements of cognitive–motor abilities by capitalising on high-fidelity motion tracking and standardised test delivery. In particular, the Global CM Score offers a robust metric for parametric analyses. While future work should address the DM ceiling effect and validate these tools in diverse populations, this approach holds significant potential for enhancing the precision and sensitivity of psychological and clinical assessment.



Academic Editor: Christos J. Bouras

Received: 25 July 2025

Revised: 7 October 2025

Accepted: 10 October 2025

Published: 16 October 2025

Citation: Craig, C.; Noble, E.; Parra, M.A.; Grealley, M.A. Advancing Cognitive–Motor Assessment: Reliability and Validity of Virtual Reality-Based Testing in Elite Athletes. *Virtual Worlds* **2025**, *4*, 46. <https://doi.org/10.3390/virtualworlds4040046>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: psychometrics; virtual reality; cognitive–motor assessment

1. Introduction

Psychometrics—the science of measuring psychological attributes like intelligence, personality, and motor skills—relies on systematic, quantifiable assessments to capture human capabilities. Early psychometric tools, such as Alfred Binet’s 1905 intelligence test, focused on cognitive faculties like memory and problem-solving [1]. However, many real-world activities, such as a surgeon performing delicate procedures or an athlete navigating dynamic environments, demand the seamless integration of cognitive processes (e.g., attention, decision-making) and motor skills (e.g., coordination, balance). Cognitive–motor assessments target this interplay, measuring how thought and movement combine to produce effective actions, distinct from isolated motor tasks like dexterity tests or power measurements [2].

Beyond performance evaluation, cognitive–motor assessments serve as sensitive indicators of brain health. Changes in cognitive–motor function can signal neurological conditions, including mild traumatic brain injury, neurodegenerative diseases, stroke, or

age-related decline. Reliable tools that track these changes enable healthcare professionals to enhance diagnostic accuracy, tailor treatment plans, and monitor disease progression, making advancements in cognitive–motor testing critical for improving clinical outcomes.

Traditional cognitive–motor assessments, however, often fall short in measurement precision and reliability. For example, the Movement Assessment Battery for Children (Movement ABC) lacks objectivity in scoring movement quality [3], while the Balance Error Scoring System (BESS) fails to capture the nuanced cognitive–motor integration required for postural control [4]. In neurodegenerative diseases, tools like the Unified Parkinson’s Disease Rating Scale (UPDRS) rely on subjective reports [5], and the Multiple Sclerosis Functional Composite (MSFC) uses basic timed tasks that oversimplify cognitive–motor interactions [6]. Similarly, concussion assessments like the Sport Concussion Assessment Tool (SCAT6) [7] and Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) offer limited granularity in measuring reaction times and errors [8]. These tools often lack advanced technologies, such as motion capture, which are essential for more precise and ecologically valid cognitive–motor evaluation.

For a test to be trusted, it must demonstrate both validity—measuring what it claims to measure—and reliability, ensuring consistent results under identical conditions [9]. Content validity ensures the test’s tasks reflect the intended construct, such as incorporating varied sensory inputs to assess balance comprehensively. Construct validity confirms that the test measures the theoretical attributes, like cognitive–motor integration, rather than unrelated factors. Traditional computerised tests improve reliability through standardised stimuli but often rely on simplistic responses (e.g., mouse clicks), limiting their ability to capture complex behaviours. In contrast, tests requiring dynamic responses, such as human movement, may enhance both validity and reliability by providing objective, precise measurement of tasks that more closely reflect real-world cognitive–motor function.

Virtual reality (VR) offers a promising solution by delivering ecologically valid cognitive–motor tests that capture real-time behavioural responses with high precision [10–12], which can resemble (i.e., verisimilitude) and account for (i.e., veridicality) real-world functioning. VR ensures consistent test administration, controlling stimuli placement and timing, and records detailed movement trajectories, reducing reliance on subjective scoring. Its flexibility to simulate diverse, realistic contexts makes VR ideal for assessing cognitive–motor functions critical to neuro-cognitive evaluation.

This study aims to evaluate the robustness of data collected from a suite of 10 VR-based cognitive–motor tests, adapted from established tasks measuring balance and gait [13,14], decision-making [15], memory [16], and manual dexterity [17]. Our primary objectives are to establish the reliability and validity of these VR assessments, ensuring they accurately capture cognitive–motor function, and confirm that their data distributions align with statistical norms observed in established psychological measures, such as IQ. We hypothesise that VR-based tests will provide a highly reliable and valid method for baselining cognitive and motor behaviours, strengthening their utility in psychological research and clinical practice.

2. Materials and Methods

2.1. VR Cognitive–Motor Tests

Details of the ten VR tests (NeuroFitXR 2022.4; INCISIV Ltd., Belfast, UK) developed to assess cognitive–motor performance can be found in Table 1. The tests were deployed on an Oculus Quest 2 headset (Meta, Platforms, Inc., San Francisco, CA, USA). The VR hardware system uses accelerometers, gyroscopes, and optical motion tracking within the headset and hand controllers (<1 mm accuracy, 6 degrees of freedom [18]) to capture head and hand movements. The accuracy and validity of this hardware for tracking human

movement and measuring balance (sway) under different visual conditions have been previously established [19]. NeuroFitXR's test environment comprised a three-dimensional virtual gymnasium where participants were positioned centrally to perform tasks. Ten tests were designed to simulate sports-related skills rather than classroom-based abilities and organised around four cognitive–motor pillars: balance and gait, decision-making, manual dexterity, and memory. These tests were designed to improve upon existing cognitive–motor assessments (e.g., Movement ABC, BESS, SCAT6) by providing objective, technology-driven measurements of performance, where faster times and higher levels of test accuracy reflect superior test performance outcomes.

Table 1. The table shows 10 different cognitive–motor tests grouped into four domains of cognitive–motor function: Balance and Gait (BG), Decision-Making (DM), Manual Dexterity (MD), and Memory (ME). Descriptions of the Cognitive–Motor Tests are provided along with the test performance metrics associated with the respective tests.

Domains	Test Name	Description	Performance Metrics
Memory (ME)	Digits Backwards (DB)	Participants recall sequences of 3, 4, and 5 digits (presented visually) in reverse order via motor response.	Time (s), accuracy (% correct)
	Matching Pairs (MP)	Participants uncover matching image pairs from a grid, simulating a memory card game.	Time (s), accuracy (% correct)
Manual Dexterity (MD)	Buzz Wire Right (BWR)	Participants guide a ring-shaped wand along a wire with the right hand, testing coordination.	Time (s), accuracy (% errors)
	Buzz Wire Left (BWL)	Mirrored version of Buzz Wire Right, using the left hand.	Time (s), accuracy (% errors)
Decision-Making (DM)	Ball Pop All (BP)	Participants pop moving spheres with handheld tools, assessing visual and motor processing.	Accuracy (% targets hit)
	Ball Pop Even (BPE)	Variant of Ball Pop All, requiring popping only even-numbered spheres (response inhibition).	Accuracy (% targets hit)

Table 1. *Cont.*

Domains	Test Name	Description	Performance Metrics
Balance and Gait (BG)	Tandem Balance Left (TBL)	Tandem stance (left foot forward) under five 10 s visual conditions (control, forward/back, diagonal, tilt, dark room).	Sway (mm, head and hand movement)
	Tandem Balance Right (TBR)	Tandem stance (right foot forward) under identical visual conditions as Tandem Balance Left.	Sway (mm, head and hand movement)
	Dual-Task Balance (DTB)	Tandem stance (dominant foot forward) using writing hand to complete a projected trail-making task (numbers 1–20).	Sway speed (mm/s),
	Tandem Walk (TW)	Six heel-to-toe walks along a 3 m virtual plank, walks 4 to 6 involve stepping over a gap (40% of height) with arms bent at elbows and hands held in spheres that move with the participant.	Time (s), accuracy (% error)

2.2. Data Collection

Data used for this population study were made available under a confidential data sharing agreement between a VR sports tech company (INCISIV Ltd.) and the University of Strathclyde. One thousand nine hundred and ninety-nine (1999) individual anonymised test sessions from 829 elite male athletes (mean age = 28.1 yrs (sd = 4.5 yrs); mean height = 1.85 m (sd = 0.27 m); mean weight = 89 kg (sd = 21.9 kg)), recorded as part of their respective teams' contractual, mandatory pre-season athletic screening, were initially analysed. A trained test administrator was present throughout the testing sessions to ensure that participants performed the tests correctly. All tests were completed without footwear on a hard, non-pliable surface. Any tests that were not performed correctly (e.g., the player walked in the balance test when they should have remained still in a tandem stance) were repeated, and the previous test data was overwritten.

From the main group of 829 athletes, 742 athletes were identified as having completed the NeuroFit XR test battery twice within a 14-day period. These participants were included in the test-retest analysis and were divided into two groups. One group of participants completed the second test within 48 h (<48 h), and the other group of participants completed the second test in more than 48 h but less than 14 days from the first one.

2.3. Data Preparation

Each Cognitive Motor domain consists of at least two tests, where test performance is measured using different performance metrics. To create a meaningful measure for each domain, it is important to combine the different metrics into a single composite domain score. To ensure the statistical validity of the composite scores, the raw data underwent a rigorous, multi-step preparation process. First, incomplete sessions and users with fewer than two complete sessions were removed, leaving a final dataset of 1946 sessions. Next, a supplementary metric, namely the Inverse Efficiency Score (IES), was calculated for

Tandem Walk, Matching Pairs, Digits Backwards, Buzzwire left, and Buzzwire right, to create a more robust performance indicator that combined speed and accuracy. All metrics were then mathematically inverted so that a higher value consistently represented better performance. To address the non-normal distributions found in the raw data, a Yeo-Johnson transformation was applied to each metric, reducing skewness and approximating a normal distribution. Finally, these transformed metrics were standardised into Z-scores (mean = 0, standard deviation = 1), and outliers (scores > ± 3.0) were removed ($n = 67$) to minimise their influence on the subsequent factor analysis.

2.4. Confirmatory Factor Analysis (CFA)

A Confirmatory Factor Analysis (CFA) was performed on 1879 sessions to validate the hypothesised relationships between the observed performance metrics (e.g., sway, accuracy, IES) and the four latent test constructs they were designed to measure Balance and Gait, Decision-Making, Manual Dexterity, and Memory. Unlike an arbitrary weighting scheme where each test component receives an equal weighting, the CFA enables the determination of optimal, data-derived factor loadings, which represent the unique contribution of each metric to its respective domain (see Table 2). This approach ensures that the calculations of composite scores are grounded in the empirical relationships within the data, providing a more statistically valid representation of an individual's cognitive-motor performance.

To determine the model fit, several established tests were used to evaluate the model's validity and its accurate representation of the data. While the Chi-Square test indicated a strong overall fit, further evaluations using the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) were also performed, further confirming an excellent fit of the model to the data (values of 0.98 and higher).

Table 2. Standardised factor loadings (data-driven weights) extracted from the CFA for the test performance metrics in each of the four cognitive motor domains. These values represent the contribution of each metric to its underlying factor. Note how the TW IES metric contributes a slightly negative loading to the factor and was subsequently excluded from the BG score calculation.

Domain	Metric	Standardised Factor Loading
BG	TBL sway	1.000
	TBR sway	1.049
	DBT sway speed	0.427
	TW IES	−0.016
DM	BP Accuracy	1.000
	BPE Accuracy	0.617
MD	BWR IES	1.000
	BWL IES	0.814
ME	DB IES	1.000
	MP IES	0.528

2.5. Calculation of Composite Scores

After performing the Confirmatory Factor Analysis (CFA), the relative factor loadings for each performance metric were used to calculate a composite score (CS) for each cognitive-motor domain. This approach ensured the weights were derived from the data itself, negating the need for arbitrary weighting schemes. The final composite score for each cognitive-motor domain was calculated as the sum of the standardised Z-scores of its

own individual constituent metrics, multiplied by the respective factor loading, with the entire sum then divided by the sum of the factor loadings (see Equation (1)):

$$CS = \frac{\sum_{n=1}^k M_n L_n}{\sum_{n=1}^k L_n} \quad (1)$$

where CS is the composite score, M_n is the standardised Z-score for the n th performance metric, L_n is the factor loading for the n th performance metric, and k is the number of metrics used in the composite score.

This normalisation step, as shown in the CS formula above, makes the final composite score comparable across domains, regardless of the number of metrics the domain contains. The only exception was the Tandem Walk (TW) metric, which was excluded from the Balance and Gait (BG) composite score due to its negligible and slightly negative factor loading. The lack of foot sensors in the TW task meant that task adherence (i.e., walking heel to toe along a virtual plank) could not be monitored, providing limited confidence that TW was a valid test. The Global Cognitive–Motor Score (Global CM) that combines all four composite scores, Balance and Gait (BG), Decision-Making (DM), Manual Dexterity (MD), and Memory (ME), was calculated as an unweighted arithmetic mean. This Global CM score provided a unified metric that reflects the overarching cognitive–motor performance.

3. Results

3.1. Distribution of Scores

The final composite scores were then assessed for normality using histograms, Q-Q plots, and three statistical tests (Shapiro–Wilk, D’Agostino–Pearson, and Kolmogorov–Smirnov) (see Figure 1). All the results from the Shapiro–Wilk test showed that the scores were normally distributed ($p > 0.05$), apart from the DM composite score ($p = 0.01$). This finding was visually confirmed by the Q-Q plots (Quantile–Quantile plots), which compare the quantiles of each dataset against the theoretical quantiles of a perfectly normal distribution. The data for three domains (BG, MD, ME) and the Global CM Score fell on a straight diagonal line, while the Q-Q plot for the DM composite score showed slight curves in the tails of the distribution, indicating a departure from normality.

3.2. Test–Retest Reliability

The test–retest reliability of the composite scores was evaluated to determine their stability over time. Given that the time interval between tests varied, the analysis was stratified into two groups based on the inter-session interval: two baseline tests conducted within 48 h of each other ($N = 543$ participants) and two baseline tests conducted more than 48 h apart (maximum 14 days) ($N = 199$ participants). This approach allows for a clearer understanding of the tool’s consistency across both short and longer timeframes.

For this analysis, a two-way mixed-effects model with a consistency definition (ICC(3,k)) was used. This model is most appropriate to assess the consistency of participant performance rankings across sessions, which is suitable for motor tasks where scores are expected to show a strong relationship rather than being identical. As the VR protocol recommends averaging two sessions to establish a stable baseline, both the single-measure reliability (ICC(3,1)) and the more clinically relevant average-measures reliability (ICC(3,k)) values are reported.

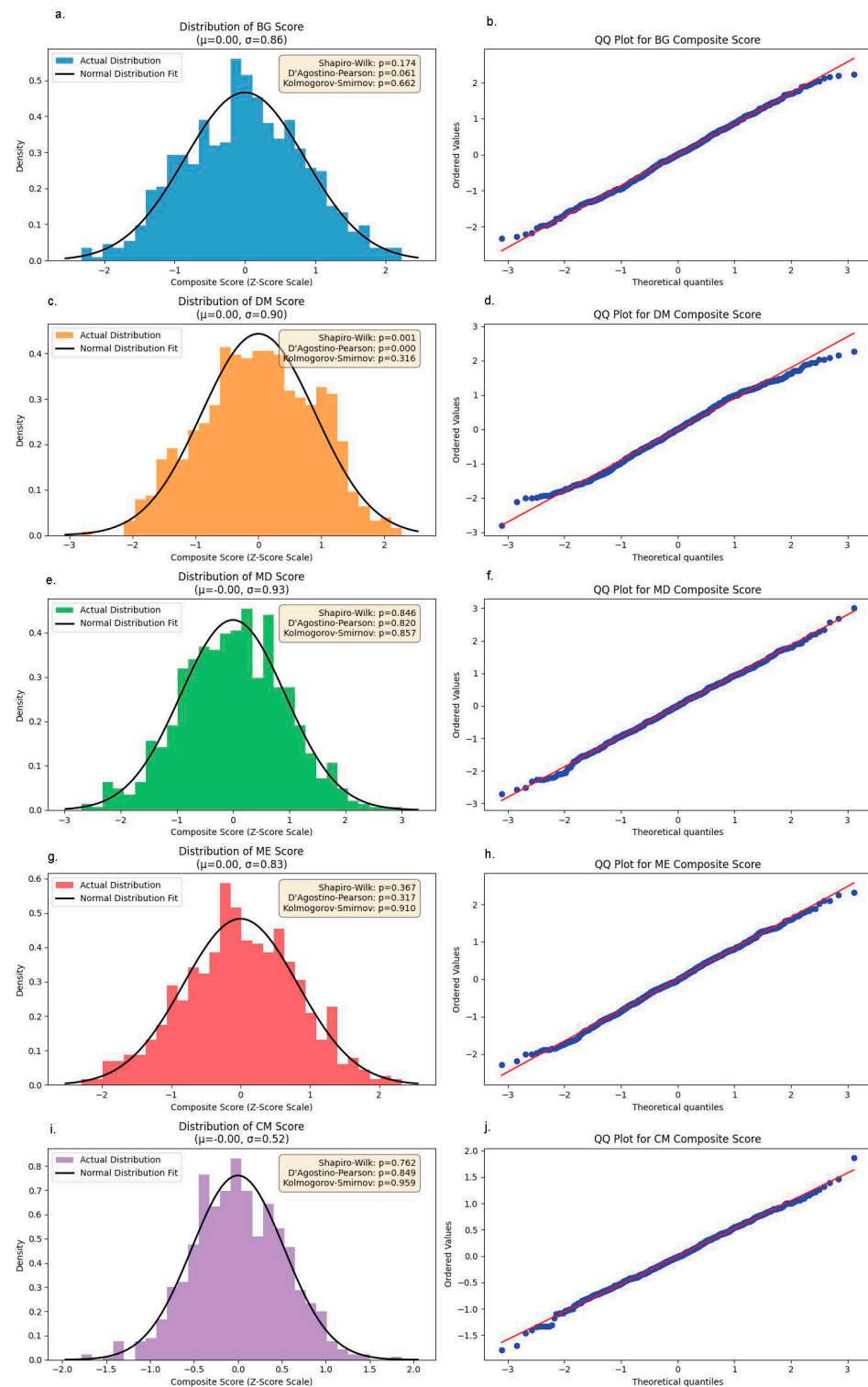


Figure 1. Histograms showing the distribution of composite scores for each of the four pillars of brain fitness along with the global score that combines all four: (a) Balance and Gait (BG score); (c) Decision-Making (DM Score); (e) Manual Dexterity (MD score); (g) Memory (ME score); and (i) Global CM Score. Scores are plotted on the x-axis and density on the y-axis. Density refers to the normalised frequency of data points per bin, such that the total area under the histogram equals 1. This enables a direct visual comparison between the recorded data's shape to the probability density function of a normal distribution (black curve). The Q-Q plots for the four cognitive–motor domains (b,d,f,h) and the Global Score (j) are also presented in the figure. The closer the data points fall to a straight, diagonal line, the more normally distributed the data is. Deviations from a straight line, such as curves or tails, suggest departures from normality. All the Q-Q plots corroborate the normal distributional patterns observed in the histograms, apart from the DM score (d).

As shown in Table 3, the composite scores demonstrated good to excellent test–retest reliability across all domains and for both time intervals. As anticipated, the average-measures reliability (ICC(3,k)) was consistently higher than the single-measure reliability (ICC(3,1)), with values ranging from moderate to excellent. The narrow 95% confidence intervals associated with these ICC values indicate a high degree of precision in our estimates. These findings provide strong empirical support for the protocol of using a two-session average, as it yields a more stable and reliable measurement for baseline assessments.

Table 3. This table shows a breakdown of the ICC(3,1) and ICC(3,k) coefficients for each of the four cognitive–motor domains and the global CM score for 742 participants who completed the baseline tests with two different durations between test sessions (<48 h (N = 1086 sessions; 543 participants) and >48 h (N = 398; 199 participants). The 95% Confidence Intervals are presented in square brackets after the ICC values. Pearson’s correlation analysis results and corresponding *p*-values are also reported for each cognitive–motor domain.

	<48 h			>48 h		
	ICC(3,1) [95% CI]	ICC(3,k) [95% CI]	Pearson’s <i>r</i> (<i>p</i> -Value)	ICC(3,1) [95% CI]	ICC(3,k) [95% CI]	Pearson’s <i>r</i> (<i>p</i> -Value)
BG	0.689 [0.64, 0.73]	0.816 [0.78, 0.84]	0.689 (<i>p</i> < 0.001)	0.701 [0.61, 0.78]	0.824 [0.76, 0.87]	0.702 (<i>p</i> < 0.001)
DM	0.576 [0.52, 0.63]	0.731 [0.68, 0.77]	0.578 (<i>p</i> < 0.001)	0.663 [0.56, 0.75]	0.797 [0.72, 0.85]	0.667 (<i>p</i> < 0.001)
MD	0.658 [0.61, 0.70]	0.794 [0.76, 0.83]	0.673 (<i>p</i> < 0.001)	0.604 [0.49, 0.70]	0.753 [0.66, 0.82]	0.625 (<i>p</i> < 0.001)
ME	0.480 [0.41, 0.54]	0.649 [0.58, 0.70]	0.482 (<i>p</i> < 0.001)	0.527 [0.40, 0.64]	0.690 [0.57, 0.78]	0.576 (<i>p</i> < 0.001)
CM	0.520 [0.46, 0.58]	0.685 [0.63, 0.73]	0.525 (<i>p</i> < 0.001)	0.620 [0.51, 0.71]	0.765 [0.67, 0.83]	0.634 (<i>p</i> < 0.001)

To provide complementary evidence of the relationship between test sessions, a Pearson’s correlation analysis was also conducted. The results revealed strong, positive, and statistically significant correlations between Session 1 and Session 2 scores for all composite measures (*p* < 0.001), further corroborating the high degree of performance consistency. While reliability remained high for the group tested more than 48 h apart, the slight attenuation in ICC values is expected when testing takes place over longer retest periods.

To further scrutinise the level of agreement between test and retest sessions, Bland–Altman plots were generated for each of the composite scores (Figure 2). These plots illustrate the difference between the two sessions against their mean, providing a visual representation of inter-session reliability. Across all measures (BG, DM, MD, ME, and the Global CM Score), the mean difference, or bias, was centred close to zero, indicating the absence of any systematic error or learning effect between the two testing sessions. The 95% limits of agreement were narrow, and the data points were randomly scattered around the mean, showing no obvious relationship between the measurement error and the mean score. This confirms that the assessments are equally reliable across the full spectrum of athlete performance. Taken together, the Bland–Altman plots provide strong evidence for the high test–retest reliability of the VR-based assessments.

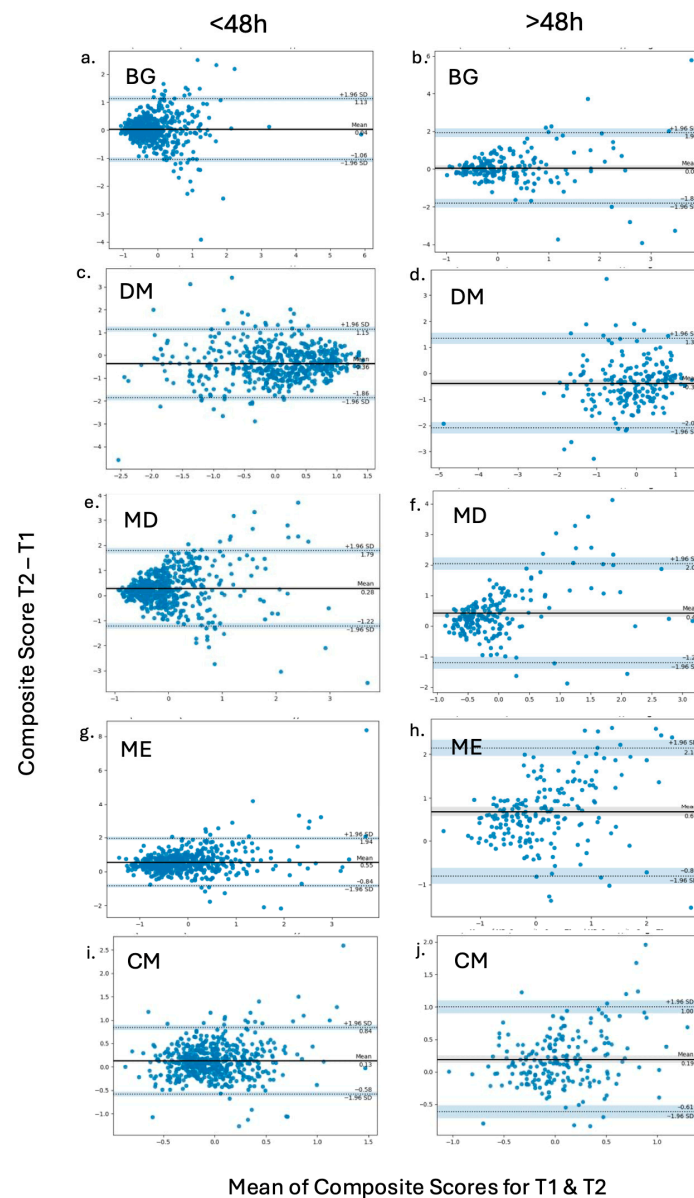


Figure 2. Bland–Altman plots illustrating the agreement between Session 1 (T1) and Session 2 (T2) for composite scores, stratified by retest interval. The plots show the difference between T2 and T1 scores (Y-axis) against the mean of T1 and T2 scores (X-axis). Panels on the left (a,c,e,g,i) represent retest intervals of less than 48 h (N = 543 participants), while panels on the right (b,d,f,h,j) represent intervals greater than 48 h (N = 199 participants). The solid horizontal line represents the mean difference (bias) between sessions, and the dashed lines indicate the 95% limits of agreement. Subplots correspond to (a,b) Balance and Gait (BG) composite score; (c,d) Decision-Making (DM) composite score; (e,f) Manual Dexterity (MD) composite score; (g,h) Memory (ME) composite score; (i,j) Global Cognitive–Motor (CM) score.

4. Discussion

This study demonstrates that cognitive–motor tests administered via VR headsets, and requiring movement-based responses, offer a robust and reliable approach to assessing cognitive–motor function. By applying a Confirmatory Factor Analysis (CFA), empirically supported factor loadings for each performance metric were derived to combine them into a single composite score for each cognitive–motor domain. A comprehensive statistical analysis of the composite scores found that three (BG, MD, ME) out of the four cognitive motor domains were normally distributed, along with the Global CM score. Despite

minor deviations at the tails for the DM test, Q-Q plots generally confirmed these findings, underscoring the need for combined visual and statistical validation [20]. When examining more closely the DM test performance metrics (i.e., accuracy), it was noted that a large proportion of athletes scored very highly on the task, creating a noticeable ceiling effect. This suggests that the DM test was too easy for this population and should be made more difficult to prevent a clustering of scores at the top end of the measurement scale.

The moderate to very good Test–Retest reliability (ICC(3,k) > 0.649 and Bland–Altman plots) for all domains across both durations indicates consistent performance across sessions. These values outperform the reliability of traditional tools like the Fugl–Meyer Assessment (FMA), Balance Error Scoring System (BESS) or Movement ABC, which often suffer from subjective scoring or limited granularity [3,4,21]. By leveraging VR’s precise motion tracking (e.g., <1 mm accuracy with Oculus Quest 2), these tests capture continuous variables—such as sway, reaction time, and movement trajectories—combining them into composite scores that offer a more nuanced view of cognitive–motor integration. This precision addresses the shortcomings of conventional assessments, which rely on binary or observational outcomes and struggle to measure complex constructs like decision-making under dynamic conditions.

VR’s ecological validity is a key strength. The VR tests used in this study were deliberately designed to mirror real-world sports skills that involve balance, anticipation, and hand–eye coordination, unlike computer-based tests that involve simplistic motor responses (i.e., button presses) [10]. For example, the Buzzwire tests, which required the participants to physically move a ring along a wire with both their right and left hands, demonstrated strong ecological validity by reflecting real-world differences in dexterity between a dominant and a non-dominant hand. A paired-samples *t*-test was conducted to compare the accuracy (100% error) of left-handed participants (*N* = 67) using their left and right hands. The results showed a statistically significant difference, with participants performing significantly better ($t_{(66)} = 2.98$; $p = 0.004$) with their dominant left hand (*M* = 92.06%, *SD* = 3.61 s) compared to their non-dominant right hand (*M* = 91.07%, *SD* = 3.90 s). Similarly, right-handed participants performed significantly better ($t_{(685)} = 10.26$, $p < 0.001$) with their dominant right hand (*M* = 91.59%, *SD* = 3.71 s) compared to their non-dominant left hand (*M* = 90.69%, *SD* = 3.98 s). These findings indicate that the task is sensitive to the established behavioural differences between dominant and non-dominant hands, suggesting it could be used as a more objective measure of manual dexterity. Future research should look to see if levels of handedness, as measured by other tests, such as the Minnesota Dexterity Test or the Edinburgh Handedness Inventory, can be detected by VR manual dexterity tasks.

Despite these strengths, limitations exist. The study’s sample—elite male athletes—limits generalisability to broader populations, such as females, older adults, or clinical groups with neurological conditions. Additionally, while VR hardware ensures precise measurements, potential confounds like user familiarity with VR or physical fatigue during testing could impact the results and were not explored in this study. Future research should validate these tests across diverse populations and investigate long-term reliability in clinical settings, such as for concussion monitoring or neurodegenerative disease tracking.

The broader implications of VR-based assessments are significant. Movement, as a biomarker of brain function, reflects the integrated activity of cognitive and motor systems [2]. VR’s ability to control stimuli and measure responses with high fidelity provides a scalable, standardised platform for detecting subtle changes in brain health, potentially revolutionising diagnostics and personalised treatment. Unlike traditional tools, such as polygraphs, which conflate physiological signals [22], or military drills, which

overlook situational factors [23], VR tests offer more immersive, context-driven assessments that balance complexity and accessibility. By overcoming the subjectivity and imprecision of conventional methods, VR-based cognitive–motor tests pave the way for more accurate, reliable, and ecologically valid evaluations of human performance and brain health.

5. Conclusions

This study highlights the transformative potential of VR in cognitive–motor assessment, offering superior precision and reliability compared to traditional methods. By leveraging VR’s high-fidelity motion tracking and immersive environments, test performance metrics, such as sway and movement response times, reveal subtle cognitive–motor differences that conventional tools often miss. The normal distribution of the composite scores, coupled with the strong Test–Retest reliability for all four domains, affirms the robustness of these assessments. This makes them ideal for standardised psychological and clinical applications, especially those requiring serial testing (e.g., post-injury assessment). VR’s ability to simulate dynamic, ecologically valid scenarios (e.g., intercepting a ball) enhances construct validity, enabling a deeper understanding of brain function using movement as a primary biomarker. Despite limitations, such as the need for broader population validation, these findings position VR as a groundbreaking tool for measuring cognitive–motor performance and advancing diagnostic precision, allowing for the personalisation of interventions in brain health research and practice.

Author Contributions: Conceptualisation and methodology, C.C.; M.A.G., E.N. and M.A.P., formal analysis, data curation, C.C. Writing—original draft, C.C. Manuscript refinement, M.A.G., E.N. and M.A.P. Supervision and project management, M.A.G., M.A.P. and E.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki. As this research involved a retrospective analysis of fully anonymised, pre-existing data, it was granted an exemption from formal ethics review. The data were collected by INCISIV Ltd. as part of mandatory, pre-season athletic screening under commercial agreements with client organisations. Under the terms of these agreements, the client organisations are responsible for obtaining player consent for the use of their anonymised data for research purposes, in line with the players’ contractual obligations. To further ensure data protection, a confidential Data Sharing Agreement was signed between INCISIV Ltd. and the University of Strathclyde.

Informed Consent Statement: Not applicable.

Data Availability Statement: Due to the commercially sensitive nature of the data, the files have not been made publicly available. Data can be made available upon request to the corresponding author.

Acknowledgments: The authors would like to thank Chloe Woods (INCISIV Ltd.), who curated the anonymised NeuroFitXR test data for validation purposes. The authors would like to extend a special thanks to Adrien Kissenpfennig of Queen’s University Belfast for his assistance with the statistical analysis.

Conflicts of Interest: Cathy Craig is a Professor at Ulster University and is also an employee at INCISIV Ltd. The other authors declare no conflicts of interest.

References

1. Binet, A. The New Methods for the Diagnosis of the Intellectual Level of Subnormals. *L’Année Psychol.* **1905**, *12*, 191–244.
2. Magill, R.A.; Anderson, D.I. *Motor Learning and Control: Concepts and Applications*, 11th ed.; McGraw-Hill Education: New York, NY, USA, 2017.
3. Henderson, S.E.; Sugden, D.A.; Barnett, A.L. *Movement Assessment Battery for Children*, 2nd ed.; Pearson: London, UK, 2007.

4. Bell, D.R.; Guskiewicz, K.M.; Clark, M.A.; Padua, D.A. Systematic review of the balance error scoring system. *Sports Health* **2011**, *3*, 287–295. [[CrossRef](#)] [[PubMed](#)]
5. Goetz, C.G.; Tilley, B.C.; Shaftman, S.R.; Stebbins, G.T.; Fahn, S.; Martinez-Martin, P.; Poewe, W.; Sampaio, C.; Stern, M.B.; Dodel, R.; et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Mov. Disord.* **2008**, *23*, 2129–2170. [[CrossRef](#)] [[PubMed](#)]
6. Cutter, G.R.; Baier, M.L.; Rudick, R.A.; Cookfair, D.L.; Fischer, J.S.; Petkau, J.; Syndulko, K.; Weinshenker, B.G.; Antel, J.P.; Confavreux, C.; et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain* **1999**, *122*, 871–882. [[CrossRef](#)] [[PubMed](#)]
7. McCrory, P.; Meeuwisse, W.; Dvořák, J.; Aubry, M.; Bailes, J.; Broglio, S.; Cantu, R.C.; Cassidy, D.; Echemendia, R.J.; Castellani, R.J.; et al. Consensus statement on concussion in sport—The 5th international conference on concussion in sport held in Berlin, October 2016. *Br. J. Sports Med.* **2017**, *51*, 838–847. [[CrossRef](#)] [[PubMed](#)]
8. Schatz, P.; Sandel, N. Sensitivity and specificity of the online version of ImpACT in high school and collegiate athletes. *Am. J. Sports Med.* **2013**, *41*, 321–326. [[CrossRef](#)] [[PubMed](#)]
9. Brown, J.D. *Testing in Language Programs*; Prentice Hall Regents: Upper Saddle River, NJ, USA, 1996.
10. Santos, F.V.; Yamaguchi, F.; Buckley, T.A.; Caccese, J.B. Virtual reality in concussion management: From lab to clinic. *J. Clin. Transl. Res.* **2020**, *5*, 148–154. [[PubMed](#)]
11. Teel, E.F.; Slobounov, S.M. Validation of a virtual reality balance module for use in clinical concussion assessment and management. *Clin. J. Sport Med.* **2015**, *25*, 144–148. [[CrossRef](#)] [[PubMed](#)]
12. Burcal, C.J.; Hagarty, A.; Grooms, D.R. Using virtual reality to treat perceptual and neurocognitive impairments after lower extremity injury. *J. Athl. Train.* **2021**, *56*, 1201–1207. [[CrossRef](#)]
13. Guskiewicz, K.M.; Ross, S.E.; Marshall, S.W. Postural stability and neuropsychological deficits after concussion in collegiate athletes. *J. Athl. Train.* **2001**, *36*, 263–273. [[PubMed](#)]
14. Slobounov, S.; Slobounov, E.; Newell, K.M. Application of virtual reality graphics in assessment of concussion. *Cyberpsychol. Behav.* **2006**, *9*, 188–191. [[CrossRef](#)] [[PubMed](#)]
15. Baker, C.S.; Cinelli, M.E. Visuomotor deficits during locomotion in previously concussed athletes 30 or more days following return to play. *Physiol. Rep.* **2014**, *2*, e12252. [[CrossRef](#)] [[PubMed](#)]
16. Collie, A.; Makdissi, M.; Maruff, P.; Bennell, K.; McCrory, P. Cognition in the days following concussion: Comparison of symptomatic versus asymptomatic athletes. *J. Neurol. Neurosurg. Psychiatry* **2006**, *77*, 241–245. [[CrossRef](#)] [[PubMed](#)]
17. Servatius, R.J.; Sproul, T.E.; Handy, J.D.; Marx, C.E.; Myers, C.E. Neurocognitive and fine motor deficits in asymptomatic adolescents during the subacute period after concussion. *J. Neurotrauma* **2018**, *35*, 1008–1014. [[CrossRef](#)] [[PubMed](#)]
18. Meta. *Meta Quest 2 Technical Specifications*; Meta: Menlo Park, CA, USA, 2023.
19. Craig, C.M.; Stafford, J.; Egorova, A.; McCabe, C.; Matthews, M. Can We Use the Oculus Quest VR Headset and Controllers to Reliably Assess Balance Stability? *Diagnostics* **2022**, *12*, 1409. [[CrossRef](#)] [[PubMed](#)]
20. Field, A. *Discovering Statistics Using IBM SPSS Statistics*, 5th ed.; SAGE Publications: London, UK, 2018.
21. Fugl-Meyer, A.R.; Jääskö, L.; Leyman, I.; Olsson, S.; Steglind, S. The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance. *Scand. J. Rehabil. Med.* **1975**, *7*, 13–31. [[CrossRef](#)] [[PubMed](#)]
22. National Research Council. *The Polygraph and Lie Detection*; National Academies Press: Washington, DC, USA, 2003. [[CrossRef](#)]
23. U.S. Army. *Field Manual 7-0: Training*; Department of the Army: Washington, DC, USA, 2019.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.