

Brief Report

Reliability of Scores Computed by a Commercial Virtual Reality System and Association with Indices of Cognitive Performance in Male Elite Rugby Players

Adrien Vachon ^{1,2,*}, Olivier Dupuy ^{2,3} , Corentin Le Moal ² and Laurent Bosquet ² 

¹ Stade Rochelais, 17000 La Rochelle, France

² Laboratory MOVE (UR 20296), Faculty of Sport Sciences (STAPS), University of Poitiers, 86000 Poitiers, France

³ School of Kinesiology and Physical Activity Sciences (EKSAP), Faculty of Medicine, University of Montreal, Montreal, QC HC3 3J7, Canada

* Correspondence: avachon@staderochelais.com; Tel.: +33-6-36325144

Abstract: Purpose: To examine the reliability of scores calculated from virtual reality (VR) games and their association with inhibitory control and cognitive flexibility in young elite rugby players. Methods: Following a familiarization session, seventeen rugby union players completed a session of a modified Stroop test and two sessions of three VR games consisting of (1) memorizing moving targets (*Tracker Master*); (2) selecting moving targets while avoiding pitfalls (*Beat Master—Never Stop*); and (3) selecting moving targets with an increasing frequency of appearance (*Beat Master—Turbo*). Results: The reliability of *Beat Master—Never Stop* was poor to moderate ($0.41 < \text{intraclass coefficient correlation [ICC]} < 0.62$; $3.2\% < \text{standard error of measurement [SEM]} < 26.1\%$), while it was good to very good for *Beat Master—Turbo* ($0.77 < \text{ICC} < 0.87$; $3.2\% < \text{SEM} < 18.2\%$). Regarding *Tracker Master*, reliability was considered as low to moderate ($0.22 < \text{ICC} < 0.60$; $2.2\% < \text{SEM} < 6.0\%$). We found strong associations between *Tracker Master* and Stroop flexibility scores ($-0.55 < r < -0.64$), as well as strong to very strong associations between *Beat Master—Never Stop* scores and the Stroop inhibition score ($0.52 < |r| < 0.84$). Conclusions: Considering their metrological properties and their association level with inhibition and flexibility, the sensibility scores of the *Beat Master—Never Stop* and *Tracker Master* games should be preferred for monitoring training load, provided at least two familiarization sessions precede them.

Keywords: virtual reality; cognitive performance; rugby union



Citation: Vachon, A.; Dupuy, O.; Le Moal, C.; Bosquet, L. Reliability of Scores Computed by a Commercial Virtual Reality System and Association with Indices of Cognitive Performance in Male Elite Rugby Players. *Appl. Sci.* **2023**, *13*, 1001. <https://doi.org/10.3390/app13021001>

Academic Editors: Jongchul Park and Samo Rauter

Received: 22 November 2022

Revised: 4 January 2023

Accepted: 8 January 2023

Published: 11 January 2023



Copyright: © 2023 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/>).

1. Introduction

Rugby union players must react efficiently and adequately in a changing and unpredictable environment. This ability requires great visual attention and efficient decision making [1]. Cognitive flexibility and inhibitory control are two high-level cognitive processes that play a significant role in this context. Cognitive flexibility represents the ability to voluntarily shift the attentional focus from one cognitive process to another. This executive function is closely related to inhibitory control, which corresponds to the ability to voluntarily inhibit an automatic response when necessary. Beyond their essential role in rugby performance, cognitive flexibility and inhibitory control have also been shown to be sensitive to fatigue [2]. Altogether, these observations suggest that specific tasks should be integrated into the training process of elite rugby players and the toolbox used to monitor internal training load or overreaching [3,4].

With the development of digital technologies, new devices such as virtual reality (VR) offer a friendly strategy to assess or develop cognitive performance. In view of their accessibility and the strong application which represents the immersive virtual environment, an increasing number of studies focus on the ergonomic aspect and assess the metrological properties of such systems in the context of the workplace [5] or elite team sport [6]. To our

knowledge, Kittel et al. [6] were among the first to assess the reliability and validity of a VR system in the context of professional sport. In their study, the authors reported better performance of elite versus sub-elite population and a better perception of environment, compared with a standard video system, concomitant with a strong relative reliability of the VR system. These results reveal that VR systems are a promising tool in the context of decision-making assessment, but their use for cognitive performance monitoring still needs to be established.

The device developed by AGON “<https://agon-league.com> (accessed on 21 November 2022)” warrants calculating several indices that are supposed to reflect cognitive performance. These scores can be used to evaluate a training intervention’s effectiveness or monitor the internal training load. However, the reliability of these scores and their association with standard measures of executive cognitive performance remains to be determined. This is the purpose of the present study, which will be implemented with young elite rugby players.

2. Methods

2.1. Participants

In total, 17 young (U21) elite rugby union players from the same Top 14 (first division of French professional rugby union) club participated in this study. The participants were considered for inclusion if they did not undergo a medical treatment known to affect cardiovascular function or cognitive performance. The final sample size was 17 players (age, 18.9 ± 0.9 years; height, 181.3 ± 6.5 cm; body mass, 91 ± 13.5 kg).

2.2. Experimental Design

Participants completed four consecutive sessions within a four-week period. The first session was dedicated to the measurement of anthropometric characteristics and familiarization with the computerized modified Stroop task. During the second session, participants performed the computerized modified Stroop task, considered as the reference score for cognitive flexibility and inhibitory control, and a familiarization exercise with the two games described thereafter. During the third and fourth sessions, participants performed these two games in order to assess the VR systems’ measurement reliability and association with cognitive performance. Therefore, each participant performed one familiarization and two testing sessions for each VR exercise. All sessions were conducted in a quiet room with constant temperature (21°C) and luminosity.

2.3. Computerized Modified Stroop Task

The computerized modified Stroop task was based on the modified Stroop color test [7]. This test included two conditions: inhibition and switching. The answers were mapped to the letters “u”, “i”, “o”, and “p” on a QWERTY keyboard, which participants used to provide their answers with the right and the left hand. The mapping remained the same throughout the task. The order was: for the right hand, “index finger—red” then “middle finger—green”, and for the left hand, “index finger—blue” then “middle finger—yellow”. The order of this response procedure was counterbalanced across participants. The first block consisted in a classic inhibition task, which requires naming the color of a color-word, the meaning of the word being incongruent with the color itself (the word BLUE written in green). In these two conditions (i.e., naming and inhibition), a fixation cross appeared for 500 ms, followed by the word for 3000 ms. The second block consisted in a switching task, which was identical to the inhibition task, except that for 25% of the trials, a square appeared instead of the fixation cross, and participants were then asked to read the color-word, instead of naming its color. The reading trials appeared randomly throughout the block. Each of the 3 blocks contained 60 trials and the screen was blank between the trials. Before each condition, participants completed practice trials: 12 for the naming condition, 12 for the inhibition condition, and 20 for the switching condition.

2.4. Virtual Reality System

The Oculus Quest (Oculus, Facebook Technologies, Irvine, CA, USA) VR headset was used to immerse the participants in cognitive stimulation. Two specific games were developed by AGON (La Rochelle, France): *Beat Master* and *Tracker Master*. During the *Beat Master* game, participants must select moving targets with different shapes or colors. A representation of the game and the playing position is shown in Figure 1. Two versions of this game were tested: the NEVER-STOP version and the TURBO version. During the NEVER-STOP version, participants were required to select moving targets while avoiding pitfalls. The TURBO version was free of pitfalls, and participants were required to select moving targets with an increasing frequency of appearance. In the *Tracker Master* condition, once the target stopped, participants were required to select the one in a different color from the others while moving. For each game, several scores were computed:

$$\text{Achievement} = (\text{True positive} + \text{True negative})^n \quad (1)$$

$$F1 \text{ score} = \frac{2}{\frac{1}{\text{Accuracy}} + \frac{1}{\text{Sensibility}}} \quad (2)$$

$$\text{Accuracy} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \quad (3)$$

$$\text{Sensibility} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \quad (4)$$

$$\text{Efficiency} = \frac{\text{Flase positive}}{(\text{True positive} + \text{False positive} + \text{False negative} + \text{True negative})} \quad (5)$$

where True positive is the number of times that an athlete selects a target that must be selected, False positive is the number of times that an athlete selects a target that should not have been selected (a mistake), False negative is the number of times that an athlete does not select a target at all (an omission); True negative is the number of times that an athlete does not select a target that should not have been selected, and n is the number of consecutive correct choices. A representation of the game is shown in Figure 1.

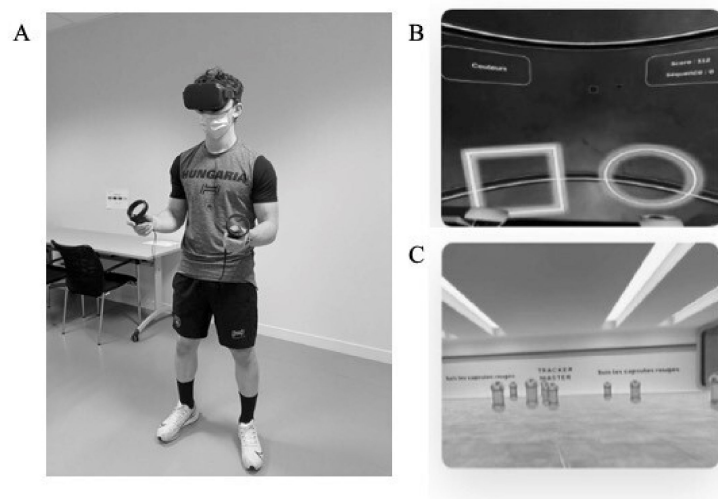


Figure 1. Player position during the virtual reality task (A) and illustration of the *Beat Master* (B) and *Tracker Master* games (C).

2.5. Statistical Analysis

Standard statistical methods were used for the calculation of means and standard deviations. A 2-way factorial analysis of variance (group $\times$ time) with repeated measures on the time factor was performed to test the null hypothesis that measures were similar

between groups and at each time point. Multiple comparisons were made with Tukey's post hoc test. The magnitude of the difference was assessed by Hedges' g (g) and considered as small ($0.2 \leq |g| < 0.5$), moderate ($0.5 \leq |g| < 0.8$), or large ($|g| \geq 0.8$) [8]. Relative and absolute reliability were assessed with the intraclass correlation coefficient (ICC; model 2.1) and the standard error of measurement (SEM). Both the ICC and SEM were computed from the breakdown two-way ANOVA (trials $\times$ subjects) with repeated measures. The ICC was considered moderate ($0.50 < \text{ICC} < 0.69$), large ($0.70 < \text{ICC} < 0.89$), or very large ($\text{ICC} > 0.90$) [9]. Standard error measurement was also used to determine the minimum difference to be considered real (MD) [10]. Pearson linear correlation (r) was used to determine the association between the VR system scores and Stroop test variables. The strength of a relationship was considered "good" with $0.69 \geq |r| \geq 0.50$, "strong" with $0.89 \geq |r| \geq 0.70$, or "very strong" with $|r| \geq 0.90$ [9]. Statistical significance was set at $p < 0.05$ for all analyses. All calculations were effectuated with Statistica 6.0 (StatSoft, Tulsa, OK, USA).

3. Results

3.1. Reliability

Reliability results are presented in Table 1. A small to moderate learning effect was observed for the scores computed during the *Beat Master—Never Stop* game, while this was large to very large for the scores computed during *Beat Master—Turbo*. The relative and absolute reliability of *Beat Master—Never Stop* was poor to moderate, while it was good to very good for the *Beat Master—Turbo* version. Regarding the *Tracker Master* game, we observed a tendency toward a moderate systematic effect for Sensibility but not for the other scores. Relative and absolute reliability were considered as low to moderate.

Table 1. Reliability of the scores provided by the virtual reality device during the games.

Parameter	Test 1 (Mean $\pm$ SD)	Test 2 (Mean $\pm$ SD)	Hedges' g (g)	ICC	SEM (%)	MD (%)
<i>Beat Master—Never Stop</i>						
Achievement (%)	55.0 $\pm$ 19.8	67.7 $\pm$ 19.6 *	0.62	0.68	26.1	44.3
F1 score (%)	91.9 $\pm$ 4.8	94.6 $\pm$ 4.7	0.55	0.61	4.3	11.1
Accuracy (%)	90.2 $\pm$ 6.2	93.6 $\pm$ 6.0	0.54	0.51	6.5	16.6
Sensibility (%)	93.8 $\pm$ 4.9	95.7 $\pm$ 3.8	0.41	0.54	3.2	8.3
Efficiency (%)	87.8 $\pm$ 7.3	91.9 $\pm$ 7.1	0.55	0.58	6.7	16.6
<i>Beat Master—Turbo</i>						
Achievement (%)	57.9 $\pm$ 16.0	74.1 $\pm$ 14.5 **	1.01	0.87	18.2	33.3
F1 score (%)	92.7 $\pm$ 3.9	96.1 $\pm$ 2.9 **	0.94	0.85	3.2	8.3
Accuracy (%)	92.1 $\pm$ 3.9	95.2 $\pm$ 4.3 **	0.67	0.78	3.2	8.3
Sensibility (%)	93.6 $\pm$ 5.2	97.2 $\pm$ 1.7 *	0.87	0.77	4.2	11.1
Efficiency (%)	88.8 $\pm$ 5.8	93.9 $\pm$ 4.5 **	0.92	0.85	4.4	11.1
<i>Tracker Master</i>						
Achievement (%)	82.8 $\pm$ 5.2	85.2 $\pm$ 4.9	0.45	0.38	6.0	13.9
F1 score (%)	82.4 $\pm$ 4.0	83.9 $\pm$ 2.1	0.45	0.41	3.6	8.3
Accuracy (%)	92.0 $\pm$ 3.5	92.1 $\pm$ 1.8	0.04	0.22	3.3	8.3
Sensibility (%)	74.7 $\pm$ 4.7	77.2 $\pm$ 3.3	0.58	0.60	5.3	11.1
Efficiency (%)	88.7 $\pm$ 2.3	89.5 $\pm$ 1.3	0.37	0.24	2.2	5.5

ICC = intraclass coefficient correlation; SEM = standard error of measurement; MD = minimum difference to be considered real; * significative difference from Test 1 with $p < 0.05$; ** significative difference from Test 1 with $p < 0.01$.

3.2. Association with Stroop Indices

The association between VR scores computed during the different games and Stroop performance is presented in Table 2. We found strong associations between VR scores computed during the *Tracker Master* game and the flexibility score of the Stroop, as well as strong to very strong associations between VR scores computed during the *Beat Master—Never Stop* game and the inhibition score of the Stroop. There was no association between the VR scores of the *Beat Master—Turbo* game and Stroop performance.

Table 2. Association between the scores provided by the virtual reality device during the games and executive performance during the computerized modified Stroop task.

	Flexibility		Inhibition	
	Mean RT	% Success	Mean RT	% Success
<i>Beat Master—Never Stop</i>				
Achievement	−0.46	−0.01	−0.52 *	0.56 *
F1 score	−0.47	0.00	−0.65 **	0.72 **
Accuracy	−0.48	0.01	−0.50 *	0.60 *
Sensibility	−0.38	−0.01	−0.82 **	0.84 **
Efficiency	−0.48	0.00	−0.65 **	0.72 **
<i>Beat Master—Turbo</i>				
Achievement	−0.19	−0.30	−0.15	0.09
F1 score	−0.31	−0.27	−0.08	0.03
Accuracy	−0.35	−0.27	0.05	−0.06
Sensibility	−0.11	−0.23	−0.40	0.26
Efficiency	−0.32	−0.27	−0.07	0.03
<i>Tracker Master</i>				
Achievement	−0.21	−0.05	0.29	−0.26
F1 score	−0.57 *	0.16	−0.03	−0.05
Accuracy	0.04	−0.01	0.19	−0.29
Sensibility	−0.64 **	0.18	−0.11	0.06
Efficiency	−0.55 *	0.16	0.05	−0.14

Mean RT: mean reaction time; * significant association with $p < 0.05$; ** significant association with $p < 0.01$.

4. Discussion

This study aimed to assess the reliability of the scores calculated by a virtual reality device during two games and their association with executive performance in young elite rugby players. The main results of this study were: the relative and absolute reliability of the scores computed during the *Tracker Master* and *Beat Master—Never Stop* games were poor to moderate, while they were considered good to very good during the *Turbo* version. Interestingly, the scores were strongly to very strongly associated with executive performance.

The association between the calculated scores and inhibition or cognitive flexibility performance of the modified Stroop task (assessed by the error rate or the reaction time) was expected, since these cognitive functions are widely used both during VR games and Stroop tasks. Therefore, participants with the highest scores during the games also had lower error rates and faster reaction times. This indicates that the use of these games during the players' preparation is likely to improve these executive functions; however, this also indicates that the calculated scores can be used to monitor internal training load and prevent overreaching.

Among the different reliability indices, the MD is essential to interpret the variation of a measure over time, which is the purpose of training load monitoring. In fact, MD represents the limit under which the observed difference is within what we expect to see

in repeated testing due to the measurement's noise. The smaller this limit, the better the capacity to interpret the variations of a measure. In our study, MD ranges from 5.5 to 44.3%. This heterogeneity must be accounted for when choosing the indices used in the follow-up. It also underlines the need to standardize the conditions of use as much as possible, whether in terms of the environment or scheduling.

When considering their reliability characteristics and their level of association with executive performance, sensibility scores measured during the *Beat Master—Never Stop* game (MD = 8.3% and $r = -0.82$) or during the *Tracker Master* game (MD = 11.1% and $r = -0.64$) represent the best compromise. The absence of a learning effect during the test–retest ($p > 0.05$ and $g < 0.58$) suggests they can be used without familiarization. Furthermore, the moderate relative reliability reported for both indices (ICC = 0.54 and 0.60 for *Beat Master—Never Stop* and *Tracker Master*, respectively) are lower than results reported with a comparable VR system assessed with elite Australian football players [6]. These differences can be explained by the lack of experience in the VR environment of the players included in our study. It suggests that VR familiarization, or acclimation, sessions are required in order to accustom players to this special environment.

5. Conclusions

The association of the different scores with inhibitory control and cognitive flexibility leads us to consider that VR offers an alternative way to assess cognitive performance. Considering their metrological properties, the sensibility scores of the *Beat Master—Never Stop* and *Tracker Master* games should be preferred for monitoring the training load, provided at least two familiarization sessions precede them in order to minimize the learning effect and to acclimate players to the VR environment. Other applications could now be tested, such as the Sensibility of these scores as criteria to determine the readiness to play after a concussion.

Author Contributions: Conceptualization, L.B. and C.L.M.; methodology, L.B. and C.L.M.; software, L.B. and C.L.M.; validation, L.B.; A.V. and C.L.M.; formal analysis, C.L.M. and O.D.; investigation, C.L.M.; resources, C.L.M.; data curation, C.L.M. and A.V.; writing—original draft preparation, L.B.; O.D. and A.V.; writing—review and editing, L.B., O.D. and A.V.; visualization, A.V.; supervision, L.B. and A.V. 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 according to the guidelines of the Declaration of Helsinki and approved by a national ethics committee for non-interventional research (IRB00012476-2021-05-02-80).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pesce, C.; Tessitore, A.; Casella, R.; Pirritano, M.; Capranica, L. Focusing of visual attention at rest and during physical exercise in soccer players. *J. Sport. Sci.* **2007**, *25*, 1259–1270. [[CrossRef](#)] [[PubMed](#)]
2. Costello, S.E.; O'Neill, B.V.; Howatson, G.; van Someren, K.; Haskell-Ramsay, C.F. Detrimental effects on executive function and mood following consecutive days of repeated high-intensity sprint interval exercise in trained male sports players. *J. Sport. Sci.* **2022**, *40*, 783–796. [[CrossRef](#)] [[PubMed](#)]
3. Dupuy, O.; Renaud, M.; Bherer, L.; Bosquet, L. Effect of Functional Overreaching on Executive Functions. *Int. J. Sport. Med.* **2010**, *31*, 617–623. [[CrossRef](#)] [[PubMed](#)]
4. Dupuy, O.; Lussier, M.; Fraser, S.; Bherer, L.; Audiffren, M.; Bosquet, L. Effect of overreaching on cognitive performance and related cardiac autonomic control: Overreaching and cognitive performance. *Scand. J. Med. Sci. Sport.* **2014**, *24*, 234–242. [[CrossRef](#)] [[PubMed](#)]
5. Caporaso, T.; Grazioso, S.; Di Gironimo, G. Development of an integrated virtual reality system with wearable sensors for ergonomic evaluation of human–robot cooperative workplaces. *Sensors* **2022**, *22*, 2413. [[CrossRef](#)] [[PubMed](#)]

6. Kittel, A.; Larkin, P.; Elsworthy, N.; Spittle, M. Using 360 virtual reality as a decision-making assessment tool in sport. *J. Sci. Med. Sport* **2019**, *22*, 1049–1053. [[CrossRef](#)] [[PubMed](#)]
7. Bohnen, N.; Twijnstra, A.; Jolles, J. Performance in the Stroop color word test in relationship to the persistence of symptoms following mild head injury. *Acta Neurol. Scand.* **1992**, *85*, 116–121. [[CrossRef](#)] [[PubMed](#)]
8. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Routledge: London, UK, 2013.
9. Munro, B.H. *Statistical Methods for Health Care Research*; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2005; Volume 1.
10. Weir, J.P. Quantifying Test-Retest Reliability Using the Intraclass Correlation Coefficient and the SEM. *J. Strength Cond. Res.* **2005**, *19*, 231. [[CrossRef](#)] [[PubMed](#)]

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.