



**HAL**  
open science

# Three Virtual Reality Environments for the Assessment of Executive Functioning Using Performance Scores and Kinematics: an Embodied and Ecological Approach to Cognition

Nicolas Ribeiro, Toinon Vigier, Jieun Han, Gyu Hyun Kwon, Hojin Choi,  
Samuel Bulteau, Yannick Prié

## ► To cite this version:

Nicolas Ribeiro, Toinon Vigier, Jieun Han, Gyu Hyun Kwon, Hojin Choi, et al.. Three Virtual Reality Environments for the Assessment of Executive Functioning Using Performance Scores and Kinematics: an Embodied and Ecological Approach to Cognition. *Cyberpsychology, Behavior, and Social Networking*, 2024, 27 (2), pp.127-134. 10.1089/cyber.2023.0314 . hal-04354788

**HAL Id: hal-04354788**

**<https://hal.science/hal-04354788>**

Submitted on 18 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

**Three Virtual Reality Environments for the Assessment of Executive Functioning Using Performance Scores and Kinematics: an Embodied and Ecological Approach to Cognition**

Nicolas Ribeiro,<sup>1</sup> Toinon Vigier,<sup>1</sup> Jieun Han,<sup>2</sup> Gyu Hyun Kwon,<sup>2</sup> Hojin Choi,<sup>3</sup> Samuel Bulteau,<sup>4</sup>  
and Yannick Prié<sup>1</sup>

<sup>1</sup>Laboratory of Digital Science of Nantes (LS2N), CNRS UMR6004, Nantes Université, Nantes,  
France

<sup>2</sup>Graduate School of Technology and Innovation Management, Hanyang University, Seoul, Korea

<sup>3</sup>Department of Neurology, College of Medicine, Hanyang University, Seoul, Korea

<sup>4</sup>CHU Nantes, Psychiatry Department , Nantes , France

**Author Note**

Nicolas ribeiro is currently at the CRP-CPO, UR 7273, UPJV, Amiens, France. We have no conflicts of interest to disclose. Correspondence concerning this article should be addressed to Yannick Prié.

Email: [yannick.prie@univ-nantes.fr](mailto:yannick.prie@univ-nantes.fr) . Samuel Bulteau and Yannick contributed equally to the research.

## Abstract

Virtual reality is an attractive technology for cognitive assessment, as it provides a more embodied experience compared to typical test situations, such as those using paper and pencil. In addition, virtual reality can immerse individuals in complex situations similar to real-life ones, thereby improving the ecological validity (i.e., face validity) of the assessment. Virtual reality also offers improved scoring of tests as it facilitates the tracking of kinematic information and the temporal tracking of activities.

This study assesses the correlation between scores on executive function assessments using standard neuropsychological tasks in paper-and-pencil format, on a tablet, and in three immersive virtual reality environments, each designed to involve specific aspects of executive function. This study also aims to assess the correlation between these performance scores and a set of kinematic measures (speed, duration, and distance traveled by the hand) collected in virtual reality.

The outcomes, including performance scores and kinematic measures, correlate both with traditional assessment methods (such as paper and pencil, and computerized 2D tests) and with each other, suggesting their potential usefulness in clinical and research contexts. The discussion focuses on the advantages of embodied, situated, and spatialized tests for cognitive assessment and the benefits of kinematic tracking in virtual reality tests for the quality of this assessment.

*Keywords:* Virtual reality, cognition, executive functioning, ecological evaluation, immersive

### **Three Virtual Reality Environments for the Assessment of Executive Functioning Using Performance Scores and Kinematics: an Embodied and Ecological Approach to Cognition**

Executive function is an umbrella term that encompasses multiple integrated components, referring to, for instance, the processes that enable an individual to plan, organize, and execute task-directed behavior<sup>1</sup>. Tests of executive functions are more predictive of decline in instrumental activities of daily living than tests of other cognitive domains<sup>2</sup>. Executive functioning performances can predict academic success<sup>3</sup>, while their deficits are correlated with social problems<sup>4</sup>. From a more clinical aspect, deficits in executive function have also been associated with a range of disorders, including attention deficit hyperactivity disorder, neurodegenerative disorders, depression, and autism<sup>5</sup>. Depression is one of the most frequent causes of loss of autonomy: major depressive disorder affects up to 15% of people over 65 years of age at home and 40% of people in institutions<sup>6, 7</sup>. It is also a good example of the consubstantiality of cognitive disorders with difficulties in coping with daily life linked to the impairment of executive functions. Impairment of executive functions is present in 60% of depressive episodes in the elderly, and has an impact on autonomy and medication taking, cooking, and driving<sup>8</sup>.

Parsons<sup>9, 10, 11</sup> suggests that virtual reality (VR) is a suitable candidate for implementing an embodied, situated, and function-based approach, as it immerses individuals in a realistic and dynamic environment resembling real-life situations. As such, VR thus offers much potential in a more ecological assessment of executive functions.

VR testing provides several benefits compared to traditional cognitive assessment methods, such as the ability to enhance the ecological validity of assessments within the consultation room<sup>9</sup>. For example, the multiple errand test which usually needs to be performed outside the consultation room in a way incompatible with day-to-day practice has been adapted to VR while retaining its validity<sup>12</sup>. Additionally, VR enables the tracking of temporal aspects of behavior and the control of the setting in which the individual is immersed during the assessment process<sup>13</sup>. However, VR tests with these advantages are not yet widely used in typical assessment; one explanation is the lack of available tools and sufficient exploration of the psychometric properties of the tests<sup>14</sup>.

In this context, we developed three immersive virtual reality environments dedicated to the assessment of executive functions and compared their relationship with each other and with previous typical assessment methods, namely, paper/pencil, or a computerized task on 2D monitor.

### **Purpose of the Present Study**

The aim of this research is to assess how performance scores and kinematic measures were obtained using three virtual reality environments to assess individuals' executive functions correlate with those of typical neuropsychological assessments of executive functions.

We expect the results of the VR executive function assessments to be consistent with the 2D paper-and-pencil and computer-based assessments. In addition, we also expect the performance and kinematics outcomes to be correlated across VR tests.

### **Method**

The study was conducted with the approval of the Ethics Committee for Non-interventional Research of Nantes University (n°15112021). It was pre-registered prior to any experimentation (OSF: 1.17605/OSF.IO/KASJ5).

### **Sample**

A sample of 60 participants, including 35 women and 25 men, was recruited through a platform called Nantes XP lab, which connects potential participants with laboratories. The ages of the participants ranged from 18 to 72 years ( $M = 43.25$ ,  $SD = 17.28$ ). They were all fluent French speakers. With the exception of one participant, all had graduated high school. They were compensated with €40 for their participation. Participants were invited to the laboratory for two separate sessions, each lasting up to one and a half hours, with a maximum of seven days between sessions to prevent significant fluctuations in cognitive states.

Inclusion criteria for the study were as follows: Participants must not have been diagnosed with cognitive impairment, have poor stereoscopic vision, epilepsy, be pregnant, have heart or balance problems, have a psychiatric history, experience discomfort in simulators, or be taking medication or substances with psychoactive effects.

Three participants were left-handed, three were partially ambidextrous and right-handed, and the remaining participants were right-handed. In terms of technological literacy, 55 participants had previous experience with VR, with 47 reporting using VR once a year or less. In terms of technological literacy, all participants used a personal computer on a weekly basis, or more frequently, and all but one used a touch phone on a weekly basis, or more frequently, ensuring that all had a comparable familiarity with computerized technology.

The sample size was chosen *a priori*, balancing practical constraints such as time, budget, and participant availability with insights from a literature review of studies on ecological VR tests (refer to the supplementary materials for the table of selected studies and participant details). Given the novel approach of the current study, we aimed to recruit as many participants as these constraints would permit. We noted that among the 38 studies we reviewed, the majority (29) had fewer participants than our selected number, yet all reported statistically significant outcomes.

### **Instrument**

The VR tasks were administered using the Meta Quest 2™ which is a stand-alone head-mounted display that allows 1832 x 1920 pixels per eye to be displayed at up to 90 Hz in a 90° field of view. Classical neuropsychological tests were administered in a quiet room, in accordance with clinical test administration standards such as the ones described in the GREFEX<sup>15</sup>.

### **Familiarization VR environment**

The familiarization environment was intended to provide the participant with prior experience of the materials and interactions present in the other three environments. It also ensured that the participant could read the instructions and clearly see a sign placed 3 m away. In this environment, the participant had to grasp a wooden tablet containing instructions and read them, sort cubes on two tables according to their color, read information on a touch screen, interact with it using the virtual hand, and visually explore the surroundings to find a painting on a wall and read aloud a series of letters.

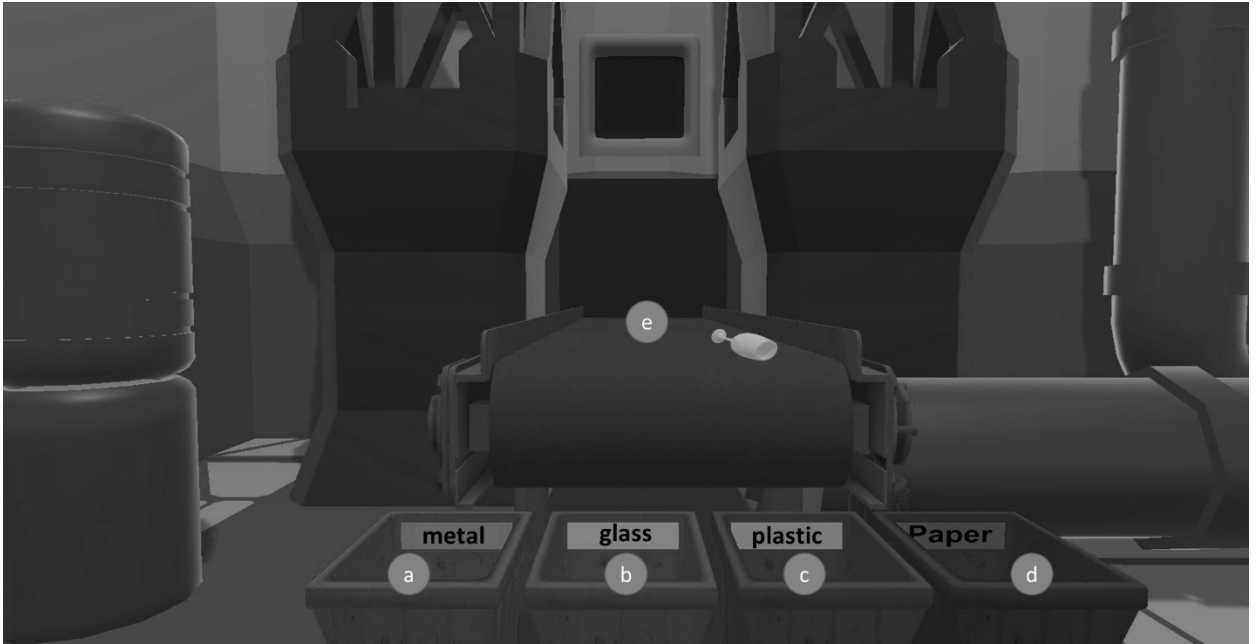
### ***Tests VR environments***

The three environments we used are adapted from proto-environments we originally designed for PC VR, where the VR headset is connected to a PC. Of the seven proposals we had, three were deemed the most promising, namely: Bel<sup>16</sup>, Mart<sup>17</sup> and Shelves<sup>18</sup>. These environments had been pre-tested in a clinical context, and we have iteratively and extensively redesigned them based on the lessons we learned from these pre-tests. We modified them to be compatible with wireless VR headsets, a Meta Quest 2 in the present study, in order to provide greater comfort for users, and we focused on usability to ensure that they were perfectly suited to real-world use.

### **Belt environment**

The Belt environment places participants in front of a conveyor belt on which they must pick up objects (Figure 1). Their task is to deposit the object they have picked up in an appropriate bin from among four bins marked: plastic, metal, glass and paper. The criterion for determining which bin to choose is the material of the object they picked up: plastic, metal, glass and paper. At the cognitive level, the participant must inhibit the color of the object (green, yellow, red and blue) in order to complete the task, as these colors coincide with those of the four bins. The position and color of the bins and objects are randomized between rounds. Belt is based on the Stroop dynamic described as mobilizing the inhibition functions in particular.

--- PLEASE PLACE FIGURE 1 ABOUT HERE ---

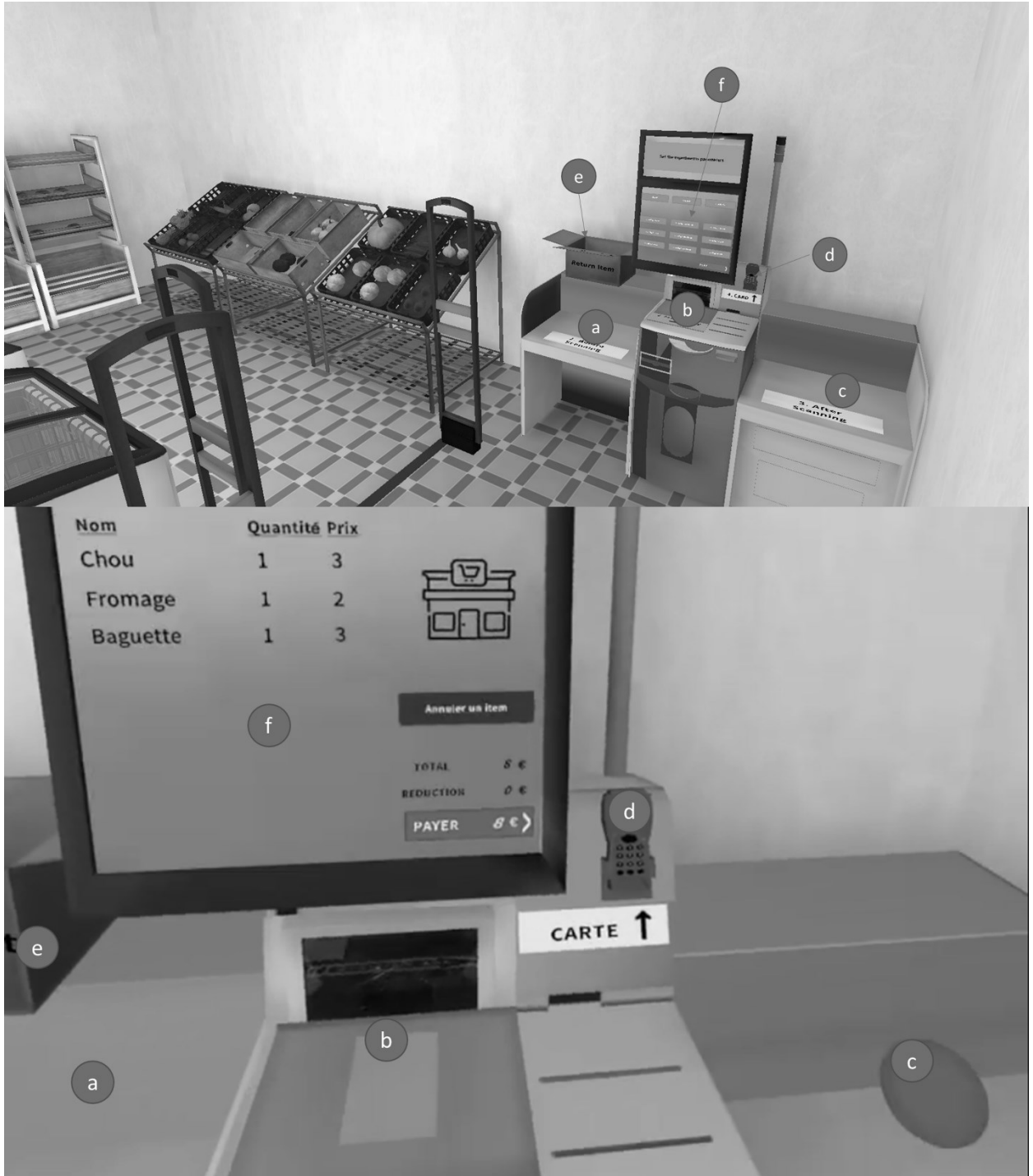


### **Mart environment**

The Mart environment immerses participants in a simulation of a supermarket. Participants are situated in front of an automatic self-checkout (Figure 2) and are informed, according to a predefined scenario, of an event that motivated their shopping such as the preparation of a picnic, and of a maximum amount that they can spend on their purchases. They are also informed of a discount code via a smartphone shown in the environment and located in their right hand. Items are placed in a first bin to the left of the self-checkout touchscreen and must pass a price detector below the touchscreen and then be placed in a second bin to the right of the touchscreen. The task of the participants in the Mart environment is to scan items, check them out, and control the cost by cancelling items depending on the money they have while keeping the most appropriate items. For instance, cancelling a frying pan in the picnic context. The participant also must remember a promotional code that they must input on a touch screen. Mart is based on the multiple errand test, which mobilizes planning in particular.

--- PLEASE PLACE FIGURE 2 ABOUT HERE ---





**Shelves environment**

The Shelves environment immerses the individual in a room with a bookshelf (Figure 3). Individuals must transfer books located on a table near the shelves to the library according to a specific sorting criterion that can be the color, size or thematic of the book. This sorting criterion varies with the

current book during the course of the task and the participant must regularly adapt to new criteria. In other words, there are three shelves, each belonging to a different person whose name is marked at the top of the shelf. Each shelf is sorted according to a different rule. The owner is indicated on the back of each book. The participant has to take a book, read the name of the owner and place it on the right shelf according to the sorting criteria. Shelves relies on the dynamics of the Wisconsin card arrangement test, which mobilizes mental flexibility and switching in particular.

--- PLEASE PLACE FIGURE 3 ABOUT HERE ---



### *Typical non-VR tests*

We selected the standard tests based on our experience in clinical neuropsychology and psychiatry and from the GREFEX cohort, which is the most widely used in evaluation consultations in France<sup>15</sup>.

### **Paper and pencil**

Typical paper-and-pencil tests for the general assessment of cognitive functions were the MOCA and the MMSE. Main paper-and-pencil tests for the assessment of executive functioning were the Stroop test, the Trail Making Test, and the Modified Card Sorting Test. Secondary tests were the digit span test, the national reading test, the Frontal Assessment Battery, and lexical and categorical verbal fluency test.

## **Computerized**

The THINC-IT tool is a computerized set of tasks aiming at assessing executive functioning<sup>19, 20</sup>. Each test gives one global score of performance. The tasks are the Choice Reaction Time (CRT), One back (NBACK), Digit Symbol Substitution Test (DSST) and the Trail Making Test (TMT). The THINC-IT has a high level of reliability and stability and acceptable level of convergent validity<sup>19, 21</sup>.

## **Design**

### ***Technical specificity of the environments***

The environments were created using the Unity engine and were engineered to maintain a framerate of 60 frames per second with a 4k resolution (2k per eye). This design was based on the research of Freiwald et al.<sup>22</sup>, who proposed that having a high framerate and resolution has a positive impact on reducing the likelihood of cyber sickness. Participants interacted with the environments using the Quest 2 controllers, which allowed for realistic object gripping, as opposed to relying on hand detection.

To further minimize the occurrence of cyber sickness, the initial exposure to VR in the familiarization environment was limited to approximately three minutes. Additionally, the duration of the test environments was approximately ten minutes, which also contributes to reducing the risk of cyber sickness<sup>23</sup>.

### ***Order of presentation***

The order of presentation of typical and VR tests has been randomized according to the following four blocks (the welcome block was always presented first).

- The first block (welcome block) included demographic questions (gender, age, education), questions on level of prior use and duration of VR, a VR familiarization session, the MOCA, the THINC-IT Cognitive Screener.
- The second block included the Belt VR environment, the Stroop and the MMSE.
- The third block included the Shelves VR environment, the modified card sorting test, the digital span and the FAB (Frontal Assessment Battery).

- The fourth block consisted of the VR Mart environment, verbal fluencies (categorical and lexical) and the National Premorbid IQ reading test.

### **Procedure**

Participants were met in a typical university consulting room and asked to perform the typical cognitive and VR tests. The VR test was administered based on block randomization and the headset was driven from a desktop to launch the application. Video content was played on the desktop to assist the participant if needed. There were no technical problems to report.

A licensed psychologist with specialization in neuropsychology administered the tests and monitored for any symptoms of cybersickness following each VR session. No participants among the 60 reported any symptoms related to cybersickness.

### **Statistical analysis**

We performed a general network analysis relying on Pearson correlations with all scores involved in the hypothesis including the paper and pencil, tablet, VR performance and kinematic scores. Correlations were corrected with a False Discovery Rate correction method<sup>24</sup>.

All score distributions and listed correlation for statistically significant tests are available in the supplementary material. In addition, the full set of data is available from the corresponding author on substantiated request.

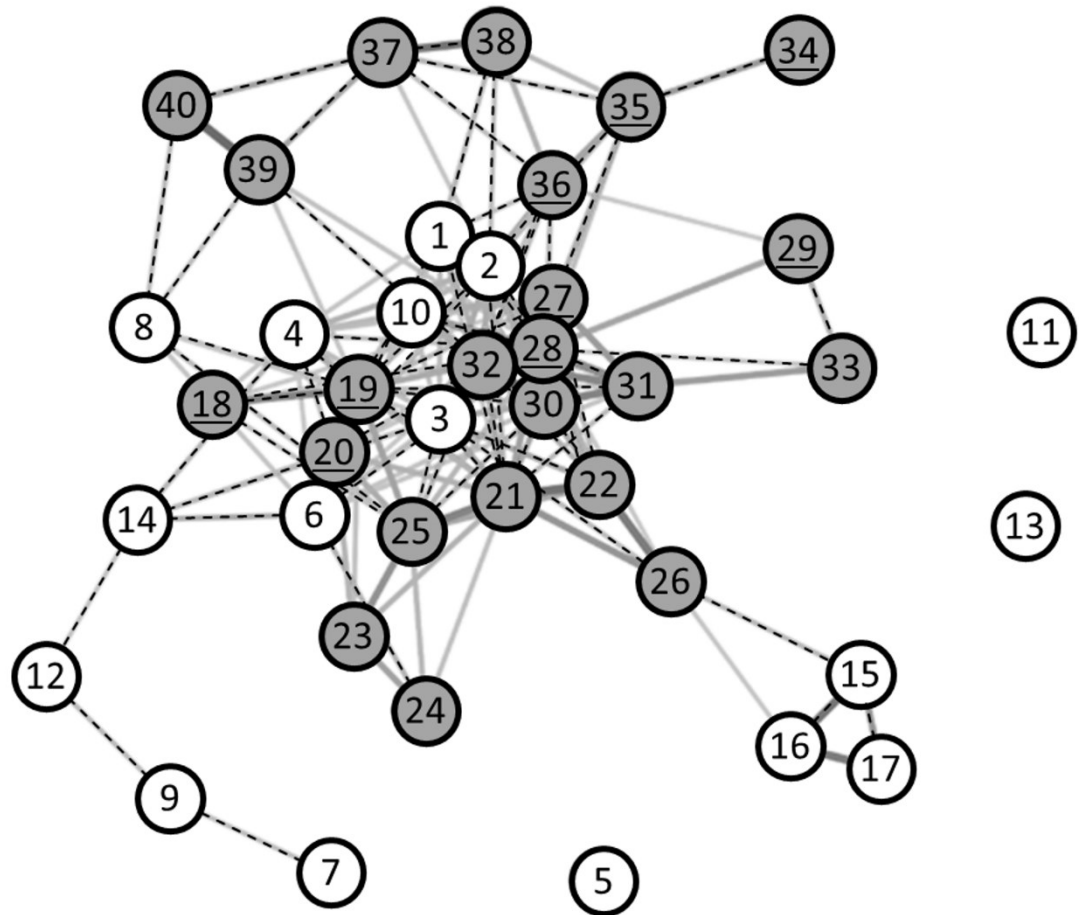
## **Results**

### ***General network analysis***

Estimated network structure illustrating how correlations are distributed across typical test results and VR tests, including scores and kinematic scores, are described in Figure 4. This network analysis most saliently shows, first, that the kinematic scores, namely speed, duration and distance, are related to other outcomes for each environment; second, that the VR test scores, namely kinematic and performance, are related to across VR tests; third, that some typical paper/pencil assessment of executive function, namely the categorical fluency score, the MOCA score, the TMT score, and the FAB score, are

related to all three VR test outcomes; fourth, that all the THINC-IT tool's outcomes are related to the VR test results; and fifth, that some of the previous paper/pencil-based outcome measures are not significantly related to each other.

--- PLEASE PLACE FIGURE 4 ABOUT HERE ---



Typical : Thinc it	Typical : Paper/pencil	MART	SHELVES	BELT
1 CRT	5 MMS	<u>18</u> Mean speed	<u>27</u> Mean speed	<u>34</u> Mean speed
2 N-Back	6 MOCA	<u>19</u> Mean duration	<u>28</u> Mean duration	<u>35</u> Mean duration
3 DSST	7 Lexical fluency	<u>20</u> Mean distance	<u>29</u> Mean distance	<u>36</u> Mean distance
4 TMT	8 Categorical fluency	21 Changes	30 Lenght pile	37 Sucess
	9 NART	22 Touch	31 All integration	38 destroy
	10 BRIEF	23 Items	32 Correct	39 Miss
	11 Stroop	24 Delta	33 Swap	40 Interference error
	12 Span direct	25 Cancel		
	13 Span inverse	26 Pay attempt		
	14 TMT			
	15 MCST category			
	16 MCST Error			
	17 MCST perseveration			

Based on the network analysis, kinematic scores correlate to other measures within the same test, and they also exhibit significant connections across different tests and test types, including both typical and VR assessments. Notably, the kinematics for the Belt test, which follows a pace dictated by the test itself, demonstrate less connectivity. This observation suggests that self-paced tests, such as Mart and Shelves, may yield richer outcomes in terms of variance shared with other cognitive tests, potentially offering greater insights into an individual's cognitive functioning. The correlation coefficients between typical and VR tests, as detailed in supplementary material, are either strong ( $r > .5$ ), moderate ( $r > .3$ ), or approaching moderate significance ( $r > .25$ ) under the typically accepted interpretation scale.

### **Discussion**

In this study, we presented three VR tests aiming at assessing executive functioning with improved ecological validity compared to previous assessment methods such as pencil-and-paper tests. The goal was to establish whether these VR tests would relate to more classical non ecological assessment methods. The three VR tests developed and evaluated in this work have a good content validity since they assess tasks directly designed from classical tests but reaching an ecological, embodied, and daily life dimension.

The results of the present study suggest concurrent validity between several outcomes of classical and VR tests as they indicate a linear relation in what is assessed with our three new VR tests and the classical tests assessing global cognitive efficiency or executive functions. Furthermore, the kinematics scores were also correlated across the three VR tests, suggesting a consistency in the potentiality of these tests to assess cognitive functioning.

Cognitive operations are intertwined with bodily information at the most fundamental level<sup>25</sup>. This suggests not only their potential to identify more direct and objective indicators through VR, but also reinforces the proposal of a neuropsychology 3.0 that is concerned with both performance and how ecological daily-life task is achieved<sup>9, 26</sup>. It is noted that these kinematics can also be obtained in real life situations with augmented reality. This proposition is reinforced by the fact that speed, duration, and distance were correlated with their counterparts in the three different VR tests, suggesting that they are

likely to be robust indicators in the sense that they retain their ability to inform functioning across the different tasks one is likely to encounter. Using augmented reality, they could thus provide an indicator of individual functioning in the most ecological situation possible, the real life setting. We also suggest more complex indicators be explored in future studies such as those listed by Clark and Riggs<sup>26</sup> which, for instance, are peak velocity, movement time, initial direction errors.

Beyond ecological validity, we believe that VR tests, such as the three environments we have presented in this article or similar tests that may be developed in the future, are a valuable addition to the clinician's toolbox. With the establishing of norms, they possibly could retain the psychometric reliability that typifies traditional tests while they already could allow for the observation of individuals in a clinical context for functional difficulties. Kinematics measures could be an early and sensible markers of cognitive-motor risk syndrome and will help in defining cognitive and behavioral trajectories in at-risk subjects and monitor longitudinally this kind of frailty.

In the present study, while typical executive functioning tests are linked with VR tests in terms of the number of connections, some tests primarily related to other cognitive domains do not show a relation to one or more VR tests. In other words, the results favor specific links between typical executive function tests and some of their counterparts for Shelves (e.g., with MCST metrics) but not for others. For example, Belt did not correlate well with the Stroop metric. However, we mainly observed this phenomenon for tests not primarily involving executive function: the span tests, which are memory tests; the NART tests, assessing lexical stock and reading skills; the lexical fluency task; and the MMS, indicative of global cognitive functioning where scores reach a ceiling effect. The lack of correlation in the latter test is more unexpected and could suggest that the Belt test fails to involve the same degree of inhibition and interference as the Stroop test, since reading is not directly involved in the Belt test once one has learned the locations of the bins and the nature of the materials. However, the Belt scores are still linked with other VR and non-VR test scores. Additional specific observations can be made on the basis of the results. First, the VR tests outcomes correlated with the MOCA but only marginally with the MMSE as suggested by the exploratory network analysis. This is consistent with the greater saturation of

executive function in the MOCA score compared to the MMSE score, and it supports the argument that the outcomes of the VR tests were indeed impacted by the participants' executive functioning performance. Second, the THINC-IT scores correlates well with the VR tests' results. Although these correlations may have been supported by technological literacy, it should be remembered that all participants had a fairly similar technological experience as all but one used touch devices on a weekly basis. It should be noted that our protocol included a familiarization environment in which all the basics of VR interaction required in the VR test environment were covered and all participants were able to complete the tests. The lack of results with specific outcomes could be related to the ceiling effect in this population with high cognitive functioning as discussed in the previous paragraph and evidenced by the score distribution (see supplementary material); consequently, clinical research are advised prior to devising further about the psychometric properties of these VR tests.

#### **Limitations and futures studies**

First, it is important to note that this study was conducted with individuals who did not have any psychological pathology. The differences in executive functioning between healthy participants may not be as important as between pathological, or between healthy and pathological ones. This may be an explanation that the strength of the adjusted correlations between VR and typical test indicators generally varies from weak to moderate, suggesting that the two methods partially but not fully align in terms of outcome variance and distribution. Yet, we continue to advocate for the ongoing evaluation of the clinical relevance of VR tests, with an emphasis on their potential to enhance our diagnostic arsenal. We believe that these two types of tests, rather than be considered adversaries, should be conceptualized as potentially synergistic tools within a clinical setting. On one hand, VR tasks, in their current form and due to the ecological nature of the tasks they offer, can already be used to observe an individual's behavior in situations that closely resemble everyday life within the consulting room. In this respect, this type of VR-based testing and environment holds significant potential for improving the clinical care and monitoring of patients at risk for cognitive-motor impairment and autonomy loss. On the other hand, VR testing cannot yet fully replace traditional testing, especially when it is necessary to assess cognitive functioning



within the scope of structuralist considerations of cognition for diagnostic purpose and classification. Also, future clinical testing of our VR environments is needed to better assess their psychometric quality notably the correlation between VR and paper-pencil test in a clinical population suffering from dysexecutive syndrome in the context of mood disorder, but also sensibility and specificity to discriminate between patients and healthy controls as well as fidelity and sensibility to change over time for longitudinal monitoring purposes.

### **Conclusion**

In total, this study evaluated how the outcomes of three VR tests aiming at assessing executive functioning and designed with ecological validity as a guide were associated with previous methods of assessment. The results of these tests, namely performance scores and kinematic indicators, are correlated with classical paper-pencil outcomes suggesting concurrent and internal validity of this innovative way to assess executive function in a more embodied and ecological way.

### **References**

1. Jurado, M. B., & Rosselli, M. (2007). The Elusive Nature of Executive Functions: A Review of our Current Understanding. *Neuropsychology Review*, *17*(3), 213–233. <https://doi.org/10.1007/s11065-007-9040-z>
2. Cahn-Weiner, D., Malloy, P., Boyle, P., Marran, M., & Salloway, S. (2000). Prediction of functional status from neuropsychological tests in community-dwelling elderly individuals. *Clinical Neuropsychology*, *14*, 187–195. [https://doi.org/10.1076/1385-4046\(200005\)14:2;1-Z;FT187](https://doi.org/10.1076/1385-4046(200005)14:2;1-Z;FT187)
3. Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, *21*(4), 327–336. <https://doi.org/10.1016/j.lindif.2011.01.007>
4. Tseng, W.-L., & Gau, S. S.-F. (2013). Executive function as a mediator in the link between attention-deficit/hyperactivity disorder and social problems. *Journal of Child Psychology and Psychiatry*, *54*(9), 996–1004. <https://doi.org/10.1111/jcpp.12072>

5. Craig, F., Margari, F., Legrottaglie, A., Palumbi, R., De Giambattista, C., & Margari, L. (2016). A review of executive function deficits in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Neuropsychiatric Disease and Treatment*, 1191. <https://doi.org/10.2147/NDT.S104620>
6. Carrière, I., Gutierrez, L. A., Pérès, K., Berr, C., Barberger-Gateau, P., Ritchie, K., & Ancelin, M. L. (2011). Late life depression and incident activity limitations: Influence of gender and symptom severity. *Journal of Affective Disorders*, 133(1–2), 42–50. <https://doi.org/10.1016/j.jad.2011.03.020>
7. Lauriola, M., Mangiacotti, A., D’Onofrio, G., Cascavilla, L., Paris, F., Ciccone, F., Greco, M., Paroni, G., Seripa, D., & Greco, A. (2018). Late-Life Depression versus Amnesic Mild Cognitive Impairment: Alzheimer’s Disease Incidence in 4 Years of Follow-Up. *Dementia and Geriatric Cognitive Disorders*, 46(3–4), 140–153. <https://doi.org/10.1159/000492489>
8. Alexopoulos, G. S. (2019). Mechanisms and treatment of late-life depression. *Translational Psychiatry*, 9(1), 188. <https://doi.org/10.1038/s41398-019-0514-6>
9. Parsons, T. D. (2015). Ecological validity in virtual reality-based neuropsychological assessment. In *Encyclopedia of Information Science and Technology, Third Edition* (pp. 1006–1015). IGI Global.
10. Parsons, T. D. (2016). Neuropsychological Assessment 3.0. In *Clinical Neuropsychology and Technology* (pp. 65–96). Springer.
11. Parsons, T. D., Carlew, A. R., Magtoto, J., & Stonecipher, K. (2017). The potential of function-led virtual environments for ecologically valid measures of executive function in experimental and clinical neuropsychology. *Neuropsychological Rehabilitation*, 27(5), 777–807.
12. Raspelli, S., Pallavicini, F., Carelli, L., Morganti, F., Pedroli, E., Cipresso, P., ... & Riva, G. (2012). Validating the Neuro VR-based virtual version of the Multiple Errands Test: preliminary results. *Presence*, 21(1), 31-42.
13. Ribeiro, N., Vigier, T., & Prié, Y. (2021). Tracking motor activity in Virtual Reality to reveal cognitive functioning: A preliminary study. *International Journal of Virtual Reality*, 21(1), 30–46. <https://doi.org/10.20870/IJVR.2021.21.1.4782>

14. Seo, K., & Ryu, H. (2018). Nothing is More Revealing than Body Movement: Measuring the Movement Kinematics in VR to Screen Dementia. *Proceedings of the Asian HCI Symposium '18 on Emerging Research Collection*, 21–24. <https://doi.org/10.1145/3205851.3205857>
15. Godefroy, O., Aithamon, B., Azouvy, P., Didic, M., Le Gall, D., & Marie, R. (2004). Groupe de reflexion sur l'évaluation des fonctions executives. *Syndromes Frontaux et Dysexecutifs*, 160, 899–909.
16. Jo, I., Han, J., Choi, H., Yannick, P., Vigier, T., Bluteau, S., & Kwon, G. (2020). Development of a Color-category Interference-based VR System for Exploring the Embodied Behavioral Patterns: Focusing on the Inhibition of Execution Function and the Phenomenology of Merleau Ponty's Perception. *Ergonomics Society of Korea*, 117–120.
17. Kim, M., Han, J., Choi, H., Yannick, P., Vigier, T., Bluteau, S., & Kwon, G. H. (2020). Implementation of Enhanced Virtual Mall for Measuring Executive Functions. *대한인간공학회 학술대회논문집*, 126–130.
18. Park, E., Han, J., Choi, H., Yannick, P., Vigier, T., Bluteau, S., & Kwon, G. H. (2020). Development of Measurement Tool for Execution Function through Behavioral Data in Virtual Bookshelves Environment: Focused on the Shifting Ability. *대한인간공학회 학술대회논문집*, 6–9.
19. Dalby, M., Annas, P., 23andMe Research Team, & Harrison, J. E. (2022). Further validation of the THINC-it tool and extension of the normative data set in a study of  $n = 10,019$  typical controls. *International Journal of Methods in Psychiatric Research*. <https://doi.org/10.1002/mpr.1922>
20. McIntyre, R. S., Best, M. W., Bowie, C. R., Carmona, N. E., Cha, D. S., Lee, Y., Subramaniapillai, M., Mansur, R. B., Barry, H., & Baune, B. T. (2017). The THINC-integrated tool (THINC-IT) screening assessment for cognitive dysfunction: Validation in patients with major depressive disorder. *The Journal of Clinical Psychiatry*, 78(7), 20938.
21. Harrison, J. E., Barry, H., Baune, B. T., Best, M. W., Bowie, C. R., Cha, D. S., Culpepper, L., Fossati, P., Greer, T. L., Harmer, C., Klag, E., Lam, R. W., Lee, Y., Mansur, R. B., Wittchen, H., & McIntyre, R. S. (2018). Stability, reliability, and validity of the THINC-it screening tool for cognitive impairment in

depression: A psychometric exploration in healthy volunteers. *International Journal of Methods in Psychiatric Research*, 27(3). <https://doi.org/10.1002/mpr.1736>

22. Freiwald, J. P., Katzakis, N., & Steinicke, F. (2018, November). Camera time warp: compensating latency in video see-through head-mounted-displays for reduced cybersickness effects. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (pp. 1-7).

23. Petri, K., Feuerstein, K., Folster, S., Bariszlovich, F., & Witte, K. (2020). Effects of Age, Gender, Familiarity with the Content, and Exposure Time on Cybersickness in Immersive Head-mounted Display Based Virtual Reality. *American Journal of Biomedical Sciences*, 12(2).

24. Benjamini, Y., & Hochberg, Y. (2000). On the Adaptive Control of the False Discovery Rate in Multiple Testing With Independent Statistics. *Journal of Educational and Behavioral Statistics*, 25(1), 60–83. <https://doi.org/10.3102/10769986025001060>

25. Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *WIREs Cognitive Science*, 4(3), 319–325. <https://doi.org/10.1002/wcs.1226>

26. Gounden, Y., Hainselin, M., Cerrotti, F., & Quaglino, V. (2017). Dynamic and Functional Approach to Human Memory in the Brain: A Clinical Neuropsychological Perspective. *Frontiers in Psychology*, 8, 688. <https://doi.org/10.3389/fpsyg.2017.00688>

27. Clark, L., & Riggs, S. (2022). VR-Based Kinematic Assessments: Examining the Effects of Task Properties on Arm Movement Kinematics. *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–8. <https://doi.org/10.1145/3491101.3519841>

**Figure 1** *Belt environment setting.*

*Note.* a, b, c, d are the four garbage cans in which the participant has to place the objects coming from (e) the Belt according to their material. Here, a yellow stemmed glass corresponding to the colour of the garbage can (a) must be placed in the blue garbage can (b), which is designated for glass material.

**Figure 2** *Mart environment setting.*

*Note.* The participant has to take items from the left side (a) of a checkout machine, scan them (b) and put them in the right side of the machine (c). Then they must try to pay for the scanned items (d) and cancel items that are too expensive for their budget in a return box (e). The machine has a touch screen (f) which is used to provide information to the participant and allow them to enter a discount code they have received on a virtual smartphone.

**Figure 3** *Shelves environment setting.*

*Note.* The participant reads a tablet containing the instructions and the context of the task. The participant then picks up a book from a table (b) and reads the name of the owner on the back of the book corresponding to one of the three Shelves (c, d, e). The name of the shelf owner is displayed at the top (f, g, h) of each. The participant must then place the book in the correct case according to the shelf owner's placement criteria.

**Figure 4** *Estimated network structure of cognitive evaluation outcomes.*

*Note.* 14 outcomes of the VR tests we proposed and 17 of traditional tests) and 9 kinematics (average speed, duration, and distance for each VR test). The kinematics labels are underlined. The estimator of the network structure is based on the correlation coefficient associated with a False Discovery Rate correction method. In a network analysis graph, the distribution of labels maximizes the proximity of correlated outcomes. Line thickness corresponds to the correlation strength. Dashed lines indicate negative correlations.