SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of Psychology

PhD Program in Psychology, Linguistics and Cognitive Neuroscience

Curriculum: Mind, Brain and Behaviour

Cycle XXXV

Exploring the Use of Virtual Environments for Neuropsychological Assessment

PhD Candidate

Luca Pieri

Registration Number

854293

Supervisor

Prof. Angelo Maravita

Co-Supervisor

Prof. Daniele Romano

Coordinator

Prof. Simona Sacchi

ACADEMIC YEAR 2021/2022

Abstract

Recently, virtual reality (VR) entered the field of neuropsychological assessment research. This technology's impact is currently discussed in the literature, as it is considered a possible promoter of a paradigm shift for neuropsychological testing. Thus, in order to contribute to this ongoing discussion, the present work has multiple purposes. First, it looks at the state-of-the-art of VR technology and its implementation in developing novel assessment tools, presenting the most important benefits of this adoption and the challenges and potential side effects to consider. Then, the core investigation is related to understanding the differences between two kinds of immersive virtual environments currently used to simulate tasks aimed at an ecological assessment of cognitive functioning: Model-Based VR and 360°-VR. To do this, two within-group experiments were carried out to compare performance and user experience-related measures during the execution of immersive VR task variants based on the two abovementioned kinds of virtual environments. The collected evidence supported the development of a third study describing a pilot VR test presented as a first-time example of a multidomain cognitive screening test based on 360°-VR. The performance of this novel-developed test in distinguishing cognitively impaired individuals from unimpaired controls was assessed, with an additional focus on investigating perceived usability and the possible occurrence of simulation sickness. Finally, a feasibility study presenting an extension of the test, including advanced data collection, such as eye-tracking and speech recognition, is offered as a possible future upgrade.

Table of Contents

List of Abbreviations……...…p.1

Introduction…… .p.2

1. Virtual Reality Technology: A Brief Introduction

1.1 Terminology, Definition and Context……..p.4

1.2 Supporting Immersion: VR Hardware……...…p.6

1.2.1 Displays and Controllers……...p.7

1.2.2 Device Tracking……..p.12

1.3 Building Worlds: VR Software……..p.15

1.3.1 Model-Based VR……p.16

1.3.2 360°-VR……...…p.18

2. Virtual Reality Technology in Neuropsychological Testing: A Systematic Review

2.1 Why VR for Neuropsychological Tests? Benefits and Challenges……..…p.21

2.2 A Short History of VR in Neuropsychological Assessment……...p.24

2.3 Methods

2.3.1 Search Strategy……...p.27

2.3.2 Study Selection……...p.27

2.3.3 Data Extraction……...p.27

2.4 Results

2.4.1 VR Tests Providing Normative Data……...p.28

2.5 Discussion

2.5.1 Cognitive Domains……..…p.30

2.5.2 Spatial Navigation…….…..p.31

2.5.3 Executive Functions……p.31

2.5.4 Attention……p.33

2.5.5 Memory……...p.35

2.5.6 Visuospatial Functions……p.36

2.5.7 Activities of Daily Living (ADL)……p.37

2.5.8 Language……p.38

2.6 Conclusion……p.38

3. Model-Based vs. 360°-VR: A Comparison of Effects on Task Performance and User Experience

3.1 Theoretical Background……p.41

3.2 Experiment One: Target Selection Task

3.2.2 Methods

3.2.2.1 Participants……..p.42

3.2.2.2 Application Design and Task Administration……p.43

3.2.2.3 Administered Questionnaires……..……p.46

3.2.2.4 Experimental Procedure……...p.47

3.2.2.5 Statistical Analysis……...p.48

3.2.3 Results

3.2.3.1 Task Performance……p.49

3.2.3.2 User-Experience…….……p.53

3.2.3.3 Correlations……..p.57

3.2.4 Discussion……...…p.60

3.3 Experiment Two: Memory Task

3.3.1 Methods

3.3.1.1 Participants……...…p.62

3.3.1.2 Application Design and Task Administration…….…….p.62

3.3.1.3 Administered Questionnaires……..……p.63

3.3.1.4 Experimental Procedure……...p.64

3.3.1.5 Statistical Analysis……...p.64

3.3.2 Results

3.3.2.1 Task Performance……p.64

3.3.2.2 User-Experience…….……p.67

3.3.2.3 Correlations……..p.71

3.3.3 Discussion and Model Comparison……p.72

4. The Mental Status 360 ° (MS-360°): Developing a Virtual Reality Test for an Ecological Screening of Cognitive Functions

4.1 Theoretical Background…….p.75

4.2 Methods

4.2.1 Participants…….p.76

4.2.2 Application Design and Task Administration…….p.77

4.2.3 Administered Questionnaires……..p.81

4.2.4 Experimental Procedure……..p.81

4.2.5 Statistical Analysis……..p.83

4.3 Results

4.3.1 Demographics…….p.83

4.3.2 Test Performance...…….p.86

4.3.3 User-Experience…….p. 92

4.4 Discussion……..p.95

5. Exploring the Use of Advanced Data Collection in 360°-VR Settings

5.1 Theoretical and Methodological Background…….p.97

5.2 Methods

5.2.1 Application Design and Task Administration…….p.97

5.2.2 Experimental Procedure……..……..p.101

5.3 Results

5.3.1 Keyword Recognition……...p.102

5.3.2 Heatmap…….p.103

5.3.3 SUS Questionnaire…….p.104

5.4 Discussion…….p.105

6. Conclusion and Future Perspectives……...p.107

References List…….……p.110

Appendix……..p.134

List of Abbreviations

1PP: First Person Perspective

3PP: Third Person Perspective

AR: Augmented Reality

BF: Bayes Factor

CGI: Computer Generated Imagery

CPU: Central Processing Unit

DoF: Degrees of Freedom

FOV: Field of View

FPS: Frames per Second

FVE: Free Visual Exploration

GPU: Graphics Processing Unit

GUI: Graphical User Interface

HC: Healthy Controls

HMD: Head-Mounted Display

IDE: Integrated Development Environment

MR: Mixed Reality

MS-360°: Mental Status 360°

POV: Point of View

RAM: Random Access Memory

SCI: Subjective Cognitive Impairment

SDK: Software Development Kit

TS: Target Selection

VE: Virtual Environment

VR: Virtual Reality

XR: Extended Reality

Introduction

During its relatively short history, the science concerning neuropsychological assessment saw the rise of many different methodological approaches for studying patients with suspected cognitive impairment (Eling, 2019). Currently, clinical observation and systematic psychometric approach remain the two core components of the evaluation process, with technologies such as brain imaging (both structural and functional) and electroencephalography (EEG) acting as complementary investigation methods from neurology. Thus, the most important advances which defined the evolution of neuropsychological assessment were related to reinventing and updating the evaluation procedures, both in terms of approaches (e.g., flexible vs fixed test batteries) and statistical methods (e.g., normative standards) (Casaletto and Heaton, 2017). More recently, the most exciting candidate considered to bring a paradigm shift in the neuropsychological testing framework is the cutting-edge technology known as virtual reality (VR) (Oker et al., 2020). This technology is not as new as one could imagine (Earnshaw et al., 1993). Still, now it almost reached a level of maturity and is widespread, allowing an evidence-based investigation of the advantages/disadvantages deriving from its implementation in the neuropsychological assessment field.

The present work is inserted in a specific declination of this research line concerning the study of the elements which different kinds of VR environments can bring in the development of novel neuropsychological tests. The theoretical part of the manuscript includes Chapters 1 and 2. The first briefly introduces the core concepts related to VR technology, focusing on the most commonly used terminology and the hardware/software solutions unique to VR. In Chapter 2, the discussion provides systematic information regarding why and how VR was implemented in neuropsychological testing during the last 20 years.

Then, the experimental part of the manuscript begins with Chapter 3, which presents two experiments aimed at investigating possible differences in users' performance and experience during the execution

of cognitive tasks implemented using two kinds of virtual environments (VEs): Model-Based VR and 360°-VR.

Chapter 4 presents the first example of how 360°-VR could be used to develop a multidomain cognitive screening test: the Mental Status 360° (MS-360°). The pilot study describes how this tool was implemented and how it detected initial cognitive impairment compared to a gold-standard instrument. Then, an example concerning what kind of advanced measures and computation methods could be added to the MS-360° is shown by the pilot study presented in Chapter 5.

Finally, Chapter 6 provides a balance concerning the evidence from the experimental studies, with a discussion regarding the possible future steps needed to advance further the knowledge related to 360°-VR environments usage in developing novel neuropsychological tests.

1. Virtual Reality Technology: A Brief Introduction[1]

1.1 Terminology, Definition and Context

The meaning of the term "virtual" changed many times over the centuries. Starting from the Latin word "virtus", which refers to an inner quality not necessarily expressed, it became later part of the computer-science world only in 1959, designating something that "is not physically existing but made to appear by software" (Harper, 2022). These two meanings share a common feature: they both imply something that does not exist yet but has the potential to be real. Then, the term "virtual reality" etymologically refers to a reality which can be *potentially* true: the judgment is ultimately given by our brain, which decides to integrate afferent multisensory signals (Riva et al., 2018). From a historical point of view, the term VR was first coined in its modern sense by Jaron Lanier in 1989 (Steuer, 1992). However, 30 years later, in its 2017 book "Dawn of the New Everything: Encounters with Reality and Virtual Reality", he noticed an excessive spread of use for this term, reporting that at least 52 different definitions of VR exist. This lack of a standardized definition was recently recognized by others (Kardong-Edgren et al., 2019), some authors proposing definitions based on a technological perspective, others focusing on the user-related experience. Technology-based definitions are more frequently adopted by developers, while user-related experience definitions are mainly adopted by behaviourists (Riva 2002; Mantovani et al., 2003; Zhou and Deng, 2009). Thus, a proper understating of what VR is can be achieved by describing the context in which this technology settles in. Currently, the term XR (i.e., extended reality) is used to encompass all the different computerized iterations generating or modifying reality, such as VR, AR (i.e., augmented reality) and MR (i.e., mixed reality). This taxonomy derives from the reality-virtuality continuum first proposed by Milgram and Kishino in 1994 and revised over the years by many authors, who however often centered their definition only on a single declination within the continuum (Rauschnabel et al., 2022).

[1] In this and subsequent Chapters, the choice of which device to present as representative of a device category was based solely on what was available at the author's laboratory. The author declares no conflict of interest.

Nevertheless, considering the most widely adopted definition of this framework, AR paradigms include all those situations in which the experience is based on a physical environment including virtual objects being superimposed and visualized using a display (e.g., a smartphone or a pair of smart glasses). On the other hand, VR paradigms refer to all those situations where the user is exposed to a fully digital world, either through a PC monitor or an immersive display. MR is instead considered an advanced form of AR, as the superimposed virtual objects can interact with the user and the real world (e.g., it is possible to grab a digital cube with the hands and place it over a real table). This modern representation of the Milgram and Kishino's reality-virtuality continuum is presented in Figure 1.

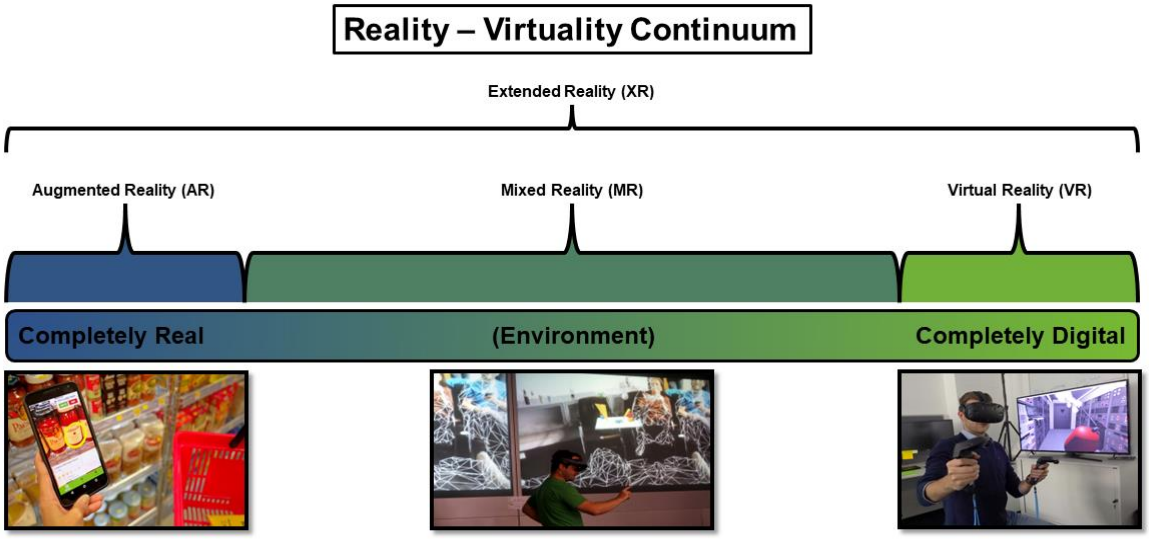


Figure 1. A modern representation of the Milgram and Kishino's continuum. The left end of the continuum defines all the environments including only physical objects, while the right end defines completely digital environments. Between the two are presented the three conventional iterations of virtual-real combinations, encompassed by the umbrella term extended reality (XR): Augmented Reality (AR), Mixed Reality (MR) and Virtual Reality (VR). For each of these technologies, an actual use-case example is presented on the lower corresponding part of the continuum. From the left to the right: an AR app providing useful information on a grocery store environment, an MR app which enables 3D environmental scanning and hologram manipulation, and a VR application for safety training. Images by (from the left to the right) Dries Buytaert (2017), Kai Kowalewski (2016), European Space Agency (ESA) (2017). All images distributed under CC BY License: https://creativecommons.org/licenses/by/4.0/.

Still, the multifaceted nature of VR cannot be fully represented by the definition given within this continuum, which remains strongly technology oriented. A more thorough classification should also include dimensions related to the user experience: that is, immersion, presence, embodiment and plausibility of the simulated environment. Briefly, the level of immersion is an objective construct, mainly related to the quality of the sensorial information provided by the VR system in giving the impression of being in the digital world. This aspect supports the emergence of three key illusions: Presence Illusion, Embodiment Illusion and Plausibility Illusion (Slater, 2009). These are all subjective constructs which concern cognitive and psychological processes at different levels. Presence refers to the illusion of being in the synthetic world (lower, perceptual level) even knowing that is not real (higher, cognitive level). Embodiment illusion refers to the perception that the virtual body is our actual body: it is achieved when the VR experience supports a visual and motor coherence between the real and the virtual body. Plausibility illusion emerges only when the events of the virtual world match the actual expectations and predictions of the user (higher, cognitive level) (e.g., virtual characters responding to the user's actions). These differences suggest that embodiment and presence illusions, being based on bottom-up elaboration, can be easily recovered when broken (e.g., restarting the application). The same is not true for the plausibility illusion which, being supported by top-end processes, it is hard to recover when it is broken. This suggests that the degree of immersion achieved relies not only on the quality of the technological system used but is also influenced by the psychological and cognitive differences existing between users (Witmer and Singer, 1998). This is a fundamental factor to consider: thus, a general definition balancing the technological and user-related features of VR experiences could be the one proposed by Riva (2022):

> "VR is simultaneously a simulative, a cognitive, and an embodied technology."

1.2 Supporting Immersion: VR Hardware

Immersion is then a key feature determining the quality of the VR simulation, as it supports user's psychological responses to VEs. The level of immersion provided by a VR system can vary along a

continuum (Fig. 2): non-immersive setups include a simple exposition (e.g., videogames on a flat desktop screen) while fully immersive setups target the highest possible number of senses. Semi-immersive setups usually include the exposition to three dimensional (3D) contents surrounding the user but do not provide the same level of interaction of fully immersive environments (Cipresso, 2018).

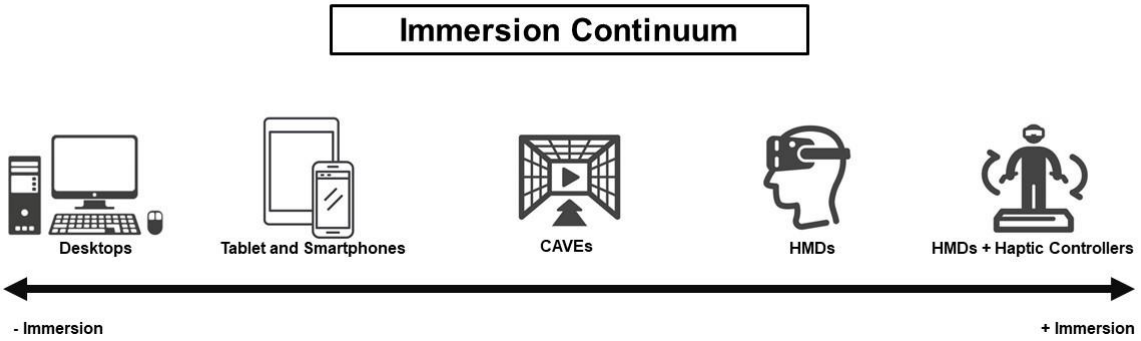


Figure 2. A graphic representation of the immersion continuum. The level of immersion provided increases going from the left to the right side of the figure. On the upper side of the line are presented some devices positioned at the corresponding level of the continuum. Desktops, tablets and smartphone supports not immersive experiences, CAVEs environments are considered semi-immersive, while HMDs setups and associated controllers provide a full-immersive setup. Abbreviations: CAVEs: Cave Automatic Virtual Environment; HMDs: Head-Mounted Displays.

1.2.1 Displays and Controllers

But what are the features of hardware systems supporting immersion? First, the core of VR experiences relies on the type of display used to visualize the digital environment. For example, flat desktop monitors provide much less immersion when compared to 360° surround screens or high-end displays embedded within head-mounted displays (HMDs). Additionally, the sensory fidelity of the experience is heavily influenced both by the visual quality and smoothness of the displayed graphics. The former relies on the colour rendering and the image clarity (i.e., the resolution) provided by the panel, while the latter is defined by the maximum frame-per-second frequency supported, (FPS, measured in Hz) also called refresh rate. This number defines the maximum processable visual

smoothness which is supported by the screen. Briefly, each 3D application is rendered by the computer hardware as a sequence of frames (frame rate) delivered subsequently to give the human eye an illusion of motion. This frame rate is defined by the design of the application and by the computational power of the rendering machine. Additionally, its maximum value is limited by the refresh rate of the display to which it is rendered. Higher framerates correspond to higher visual smoothness and to improved immersion. For VR is crucial to target 30 or better 60 FPS for non-immersive applications, and at least 90 FPS for semi-immersive/immersive applications. This is crucial not only for providing a smooth experience but also for limiting the onset of possible discomforts (Jerdan et al., 2018).

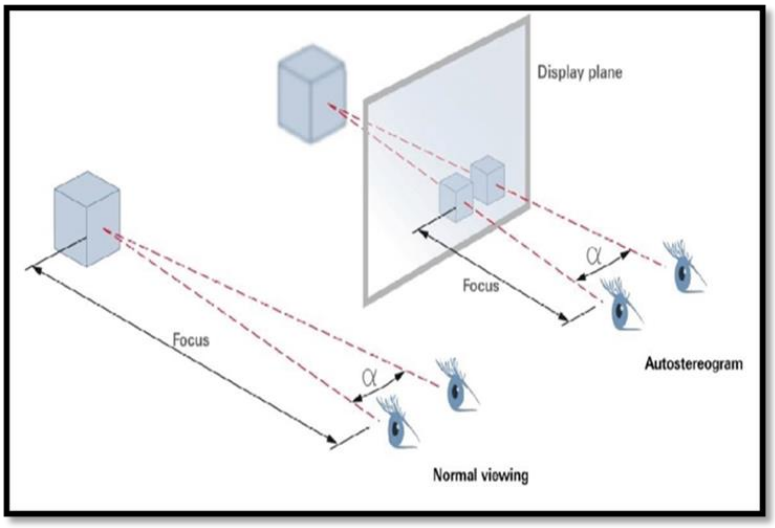
A separate case is represented by full immersive setups which include wearing an HMD (Fig. 3).



Figure 3. An example of high-end HMD, the Varjo VR-3 (https://varjo.com/products/vr-3/). Picture taken at MiBTec (https://www.mibtec.it/en/), University of Milano-Bicocca, Milan, Italy.

In this case, the display is positioned in front of the user's eyes, with a pair of lenses applying the necessary distortion to set the perceived field-of-view (FOV, measured in degrees). This parameter defines how much of the virtual world can be seen by the user inside the HMD, and it is strongly

related to immersion: wider FOVs are indeed generally associated with improved immersion (Lin, 2002). Moreover, while for non-immersive setups the 3D applications run in a 2D display thanks to the rendering process operated by the dedicated hardware, immersive displays require something quite different. Specifically, in this case the three-dimensionality of the VE is provided by a stereoscopic illusion (Stark, 1997). This is a technique used to mimic the human binocular vision (Fig. 4a). The visual system receives two different angled views of the same object, one from the right eye and one from the left eye. Then, these images are slightly different because they derive from distinct perspectives: however, the brain gather these data together in a single image to provide additional information about distance to an object, that is, depth. Within HMDs this process is simulated splitting the displayed content into two slightly different rendered frames, one for each eye (Fig. 4b).



(a) Binocular Vision



(b) HMD Stereoscopic Rendering

Figure 4. (a) A graphic representation of how human binocular vision works for processing real (left) and virtual (right) objects (Garcia et al., 2019). (b) An example of how stereoscopic VR is rendered inside the HMDs (Parisi, 2015).

Another fundamental characteristic sustaining immersion in VR environments is the possibility to control the point-of-view (POV) in which the scene is seen. Usually, first person perspectives (1PP) are considered more immersive (Denisova and Cairns, 2015) than third person ones (3PP). Most of the important interactions with the VE are provided using input interfaces, such as keyboard, mouse, gamepad or tracked controllers (Fig. 5a). Some of these can also deliver haptics sensations, which include vibration, force feedback, touch, and motion (Wang et al., 2019). These devices are fundamental also because they allow to interact with virtual objects and enable the exploration of the digital environment through movement. These two dimensions can be implemented very differently

according to whether the VR system is immersive or not. In the second case, the interaction with the virtual world is usually provided pressing buttons or moving joysticks on an interface such as a mouse, a keyboard, or a gamepad. In this context, haptics devices such as haptic data gloves can improve the quality of the interface, as they also provide vibrotactile feedbacks to the user (Fig. 5b).

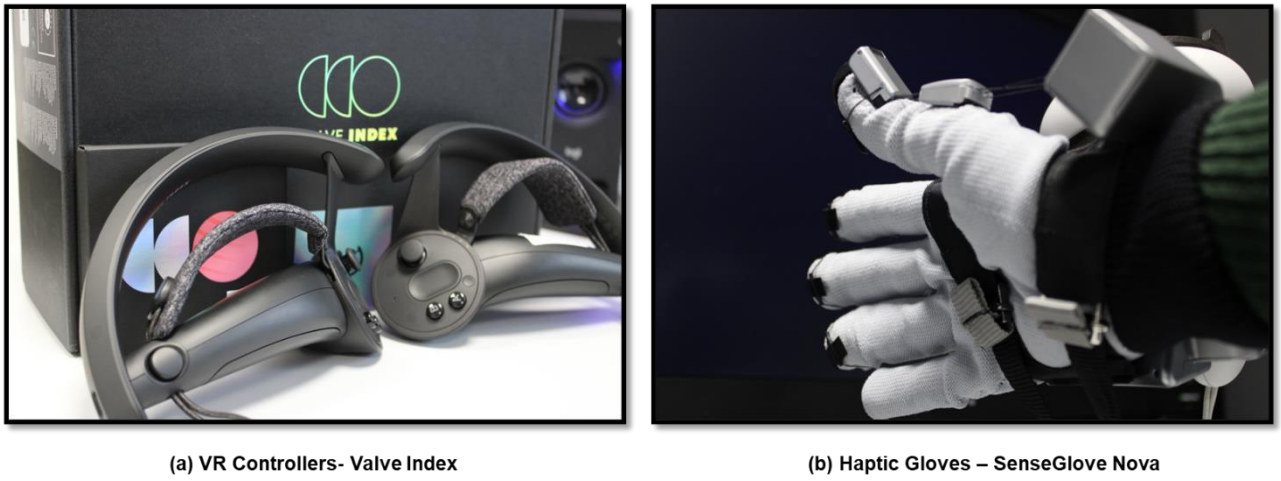


Figure 5. (a) An example of VR tracked controllers, also called wands. These are the Steam Index controllers (https://www.valvesoftware.com/en/index/controllers). (b) An example of haptic gloves. These are the SenseGlove Nova (https://www.senseglove.com/product/nova/) Picture taken by myself at MiBTec (https://www.mibtec.it/en/), University of Milano-Bicocca, Milan, Italy.

However, the most complex interactions with a virtual world can be experienced only with immersive VR devices. Indeed, in those cases the user's body become a controller, with important implications concerning how bodily sensations and expectations are elaborated in the brain (Riva et al., 2019). This is possible because HMDs and associated controllers, also called wands, include different sensors which enable accurate tracking of user's movements. For example, the movements of the user's head can be registered by the HMDs and used as input to update the POV in which the scene is seen. This dynamic control of the viewpoint is the core of immersive VR experiences. Then, the wands can be used for reaching, grasping, and manipulating virtual objects. More recently this interaction became viable also using real hands (Buckingham, 2021). Additionally, these controllers enable user's movements within the virtual world, either through the inputs on a joystick (i.e., continuous movement) or with a pointing-based paradigm called teleport (Fig. 6). The latter is usually

preferred as it can lower the chances of occurring simulation sickness symptoms (Printhul et al., 2021).

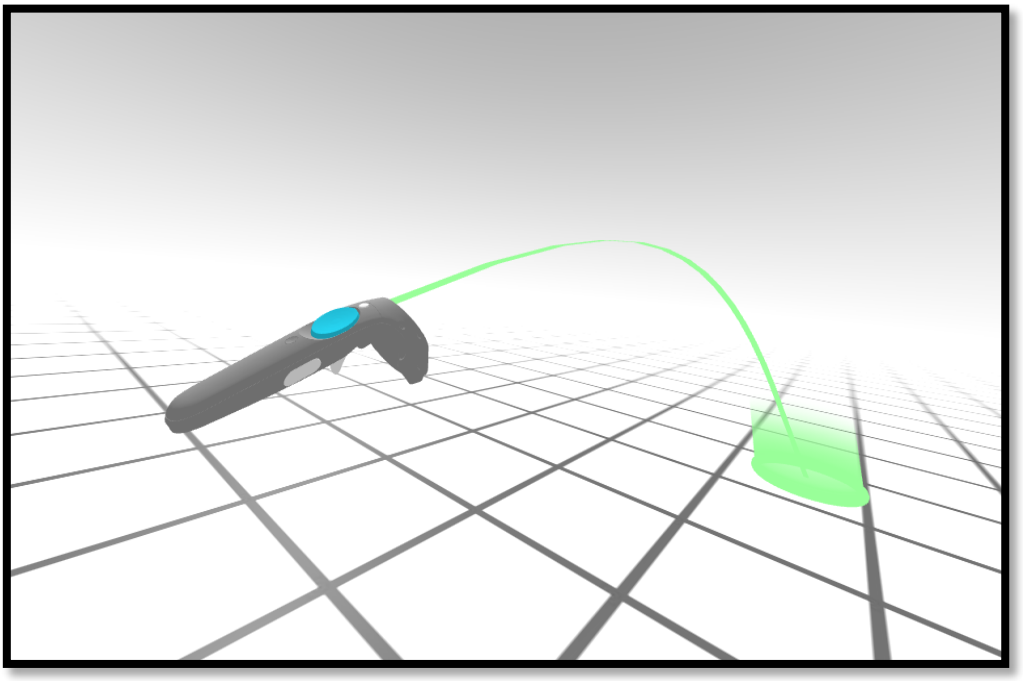


Figure 6. An example of teleporting in VR environments. The user can point and then select a portion of the virtual world to instantly being transported in that exact location. Usually, valid targets are labelled in green, while the invalid ones are labelled in red (Serrano, 2019). Image distributed under CC BY License: https://creativecommons.org/licenses/by/4.0/.

1.2.2 Device Tracking

However, users can also explore the VE through naturalistic movement. In fact, immersive VR systems always requires first the users to signal the physical borders of their playground area, in order to define the boundaries of the virtual experience, but also for safety purposes (Fig. 7). Whitin these limits the users is free to walk around as they would do in real life. Still, the physical game area often covers only a small portion of the VE: that is why joystick continuous movement and teleporting usually need to be implemented.

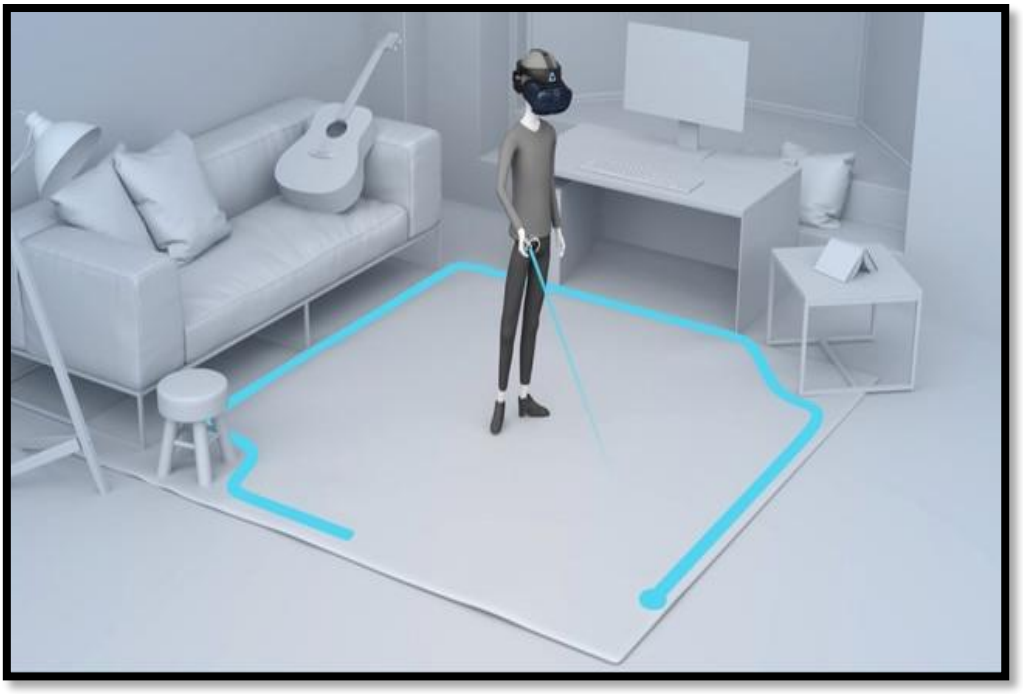


Figure 6. An infographic used for guiding the users to setup their tracking space (HTC Corporation, 2022).

But how HMDs are tracked within the physical environment? They usually integrate one out of two viable solutions: inside-out tracking and outside-in tracking (Fig. 8). Briefly, the first is performed by cameras positioned on the frontal part of the HMDs. They scan the physical environment and use the included sensors to obtain metrics data. This information is then utilized to calculate the position and the rotation of the HMD and wand controllers. These solutions are usually integrated in stand-alone HMDs which, namely, do not need external devices such as trackers or desktops to run VR applications: the computation is all done by the hardware integrated in the HMDs. On the contrary, outside-in tracking is performed by external infrared emitters whose signals are identified by detectors and cameras positioned on the HMDs. These solutions are instead generally adopted by PC-tethered HMDs, which requires a powerful desktop or laptop for running VR applications. Consequently, when compared to stand-alone devices, these setups are not so portable, but can run more complex applications and provide an overall better tracking.

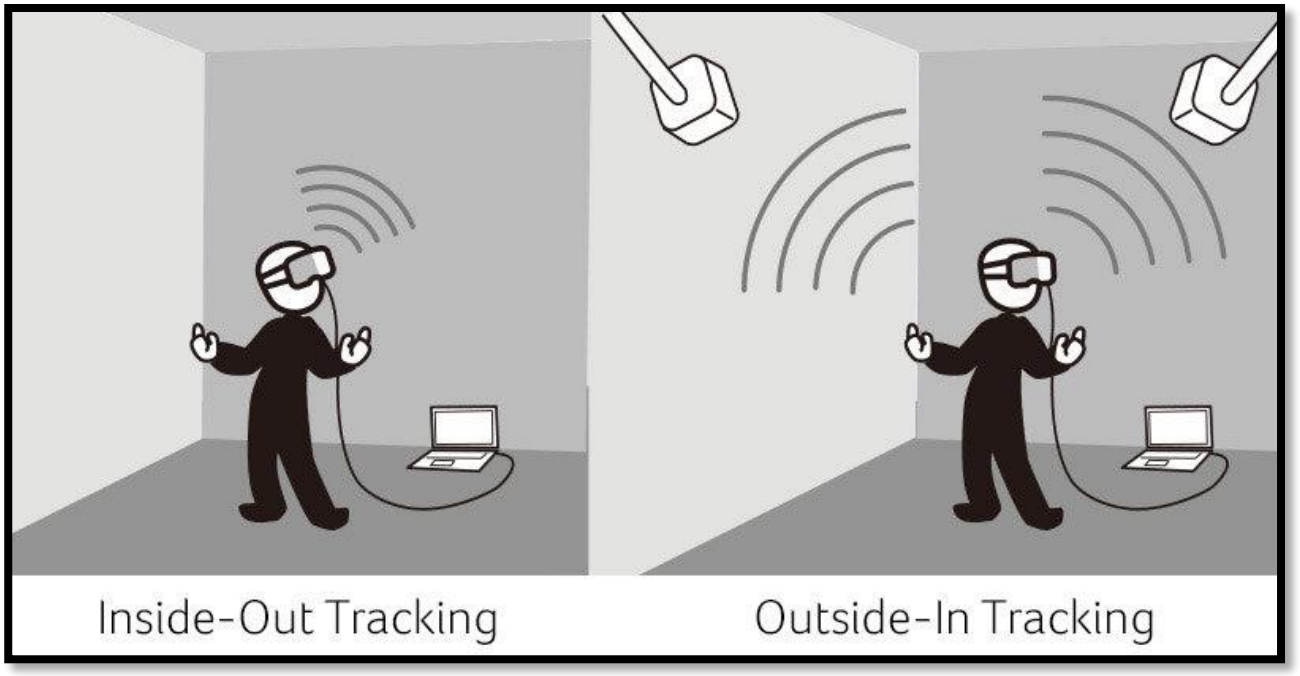


Figure 7. Inside-out versus Outside-in tracking for HMDs (Acer, 2019).

Summing up, freedom of movement plays a fundamental piece in supporting immersion. Given all the previous considerations concerning tracking, it is still possible to add a further classification. Immersive VR experiences can support two different coordinates system for tracking, implemented both by design or specified by hardware limitations: 3 or 6 Degrees of Freedom (DoF). The first allows only rotational head tracking: that is, the HMDs detects the side users are facing and follows the movements of their head allowing to turn in the virtual space while remaining physically still in one point. Thus, in this case users can move inside the virtual world only using continuous movement or teleporting, but physical moving (e.g., walking forward) is not processed by the VR application. In contrast, 6-DoF setups allow also to track the movements of the HMDs around the set play area. Thus, this tracking system enables the user to walk inside VEs, and to look at objects from all angles. Consequently, 6-DoF setups provide a higher level of immersion, as they can simulate users' movements in a more naturalistic fashion. An infographic resuming 3-DoF and 6-DoF tracking is presented in Figure 9.

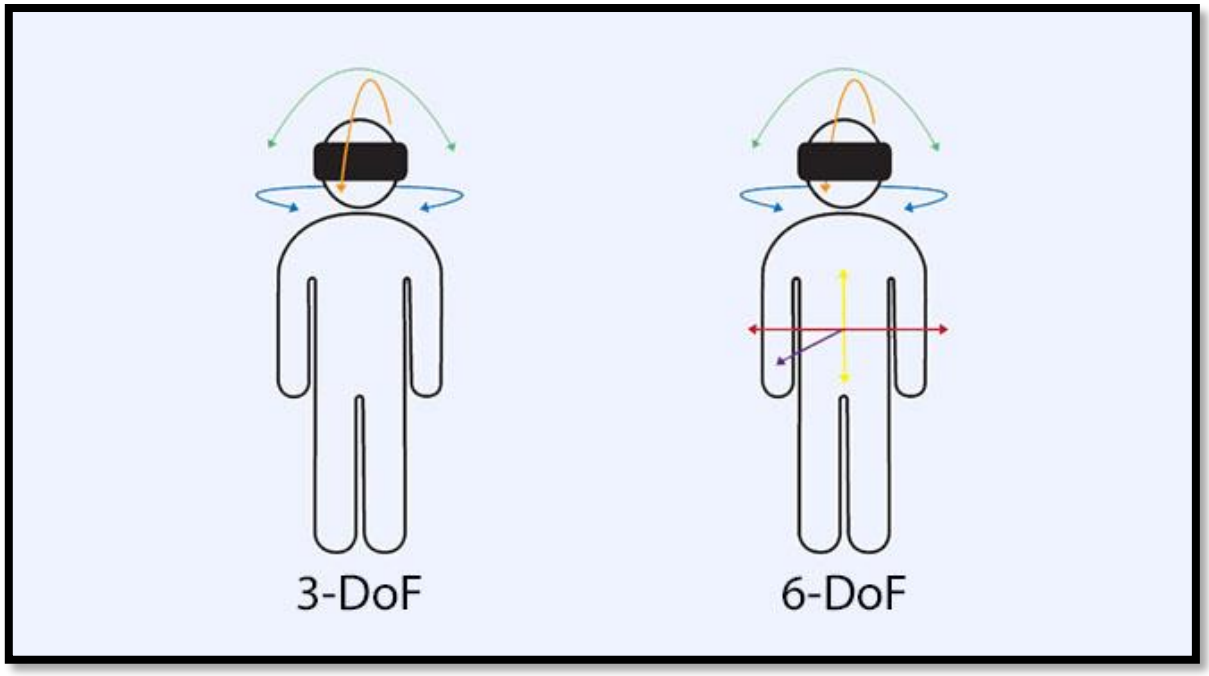


Figure 8. Viewing angles and the degrees of freedom for 3-DoF and 6-DoF tracking (Barnard, 2019).

1.3 Building Worlds: VR Software

VR environments consist in 3D scenes simulating real-life scenarios or, on the contrary, representing virtual creations which do not follow the rules of the real world (e.g., non-Euclidean geometry). The design of each single application is indeed guided by the ultimate purpose of the experience: it may be purely entertainment-based (i.e., multimedia consumption, videogames) or define a new intervention method in a specific scientific field (i.e., serious applications). Regardless the nature of their application, VR environments are created using dedicated software (e.g., Blender, 2018) implying the use of computer graphics to elaborate 3D items. Moreover, in order to generate a convincing experience, it is fundamental to integrate these objects with a system handling physics interaction, animations, and environmental sounds. These components are usually provided by game engines (e.g., Unity, 2005) a kind of software which allows to import virtual objects and program their behaviour to create a living digital world. In order to do this, game engines require an integrated development environment (IDE) supporting specific programming languages, such as C++ or C#. This allows the complete control over the behaviour of the VR scene trough scripting and handles the

creation of the actual executable file which will launch the application. Moreover, hardware manufactures usually provide software development kits (SDKs) which enables and facilitate the integration of different VR devices within a specific application.

The graphical nature of VR environments is not always the same. Until recently, it was predominantly constituted by computer-generated imagery (CGI), defining the so-called model-based VR. However, in the last years a different kind of graphic environments started to emerge. These are based on spheric image or video captures of real-world scenarios (360°-VR). They are taken using specifically designed omnidirectional cameras and then computed to give a 360° view of the photographed environment from the POV where the camera was placed during the recording.

1.3.1 Model-Based VR

3D objects are usually created using software integrating a graphical user interface (GUI), without direct interaction with the complex mathematical operation needed to compute the geometry. The process starts with modeling object's wireframe and polygons. That is, the geometrical structure is first defined by various points and lines connecting those points (i.e., wireframing). Then, the wireframe's sides are filled by faces to create a solid (i.e., polygonal modeling). The object is then textured, which means that images are applied to the surface to provide a more complex appearance (Fig. 10).

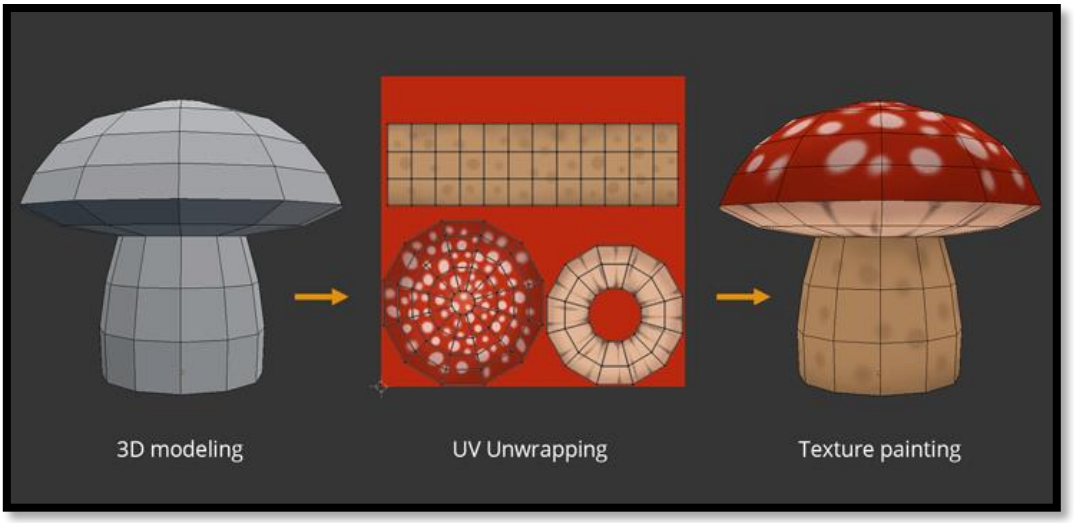


Figure 9. An example of 3D object modeling and texturing (Muttaqien, 2017).

These objects are then usually imported to game engines as graphic assets, which will be used to populate the virtual scene. Afterwards, during application runtime these objects are rendered in the display and can be visualized by the user either through a non-immersive or immersive setup. The graphic quality of the rendered model-based scenarios can vary greatly on the base of two components: the complexity of the 3D models used for building the environment and the lighting effects. Generally, immersive model-based VR scenarios are less visually complex if compared to non-immersive ones which, on the contrary, can more easily reach photorealistic fidelity on modern consumer hardware (Chalmers and Ferko al., 2010). Notably, the main reason is related to the heavy additional computational power which immersive VR require to render and update stereoscopic views of the environment based on user's dynamic POV.

For what concerns user tracking, model-based VR scenarios support both 3 and 6 DoF exploration. The virtual world coordinates are indeed fully controlled by the settings defined within the game engine: for example, depending on the purpose of the VR application it is possible to integrate 6-DoF with continuous movement, or 3-DoF with still positioning and teleporting.

1.3.2 360°-VR

Model-based environments represent the most generic form of graphic content displayed in current VR applications. However, recently a new kind of virtual scenarios emerged and began to gain the attention of researchers coming from different fields: 360°-VR. These environments are built starting from spheric photo or video captures of real-world scenes. The launch of consumer-oriented and user-friendly omnidirectional cameras (Fig. 11) contributed to the widespread diffusion of this new kind of media which, additionally, can be experienced both trough immersive and non-immersive setups.



Figure 10. An example of omnidirectional camera used for capturing 360°-VR media. This is the Insta360 One RS 1-Inch Edition (https://www.insta360.com/it/product/insta360-oners/1inch-360), available at MibTec labs where the photo has been taken (https://www.mibtec.it/en/), University of Milano-Bicocca, Milan, Italy.

First, depending on the number of camera lenses they integrate, these devices simultaneously take video or photo captures of a real-environment scenes from different perspectives. For example, the camera presented in Figure 11 takes two 180° captures, one from the front and one from back of its central axis. Then, these images are fused together to generate an equirectangular projection of the captured scene, a process call stitching (Fig. 12a). Within game engines this projection is usually

imported and then rendered in the inner surface of a sphere, with its centre corresponding to the location where the camera was put when recording. This position also matches the user's POV. Finally, the resulting media can be experienced trough a PC monitor controlling the POV with the mouse, or through an HMD with rotational tracking (Fig. 12b).

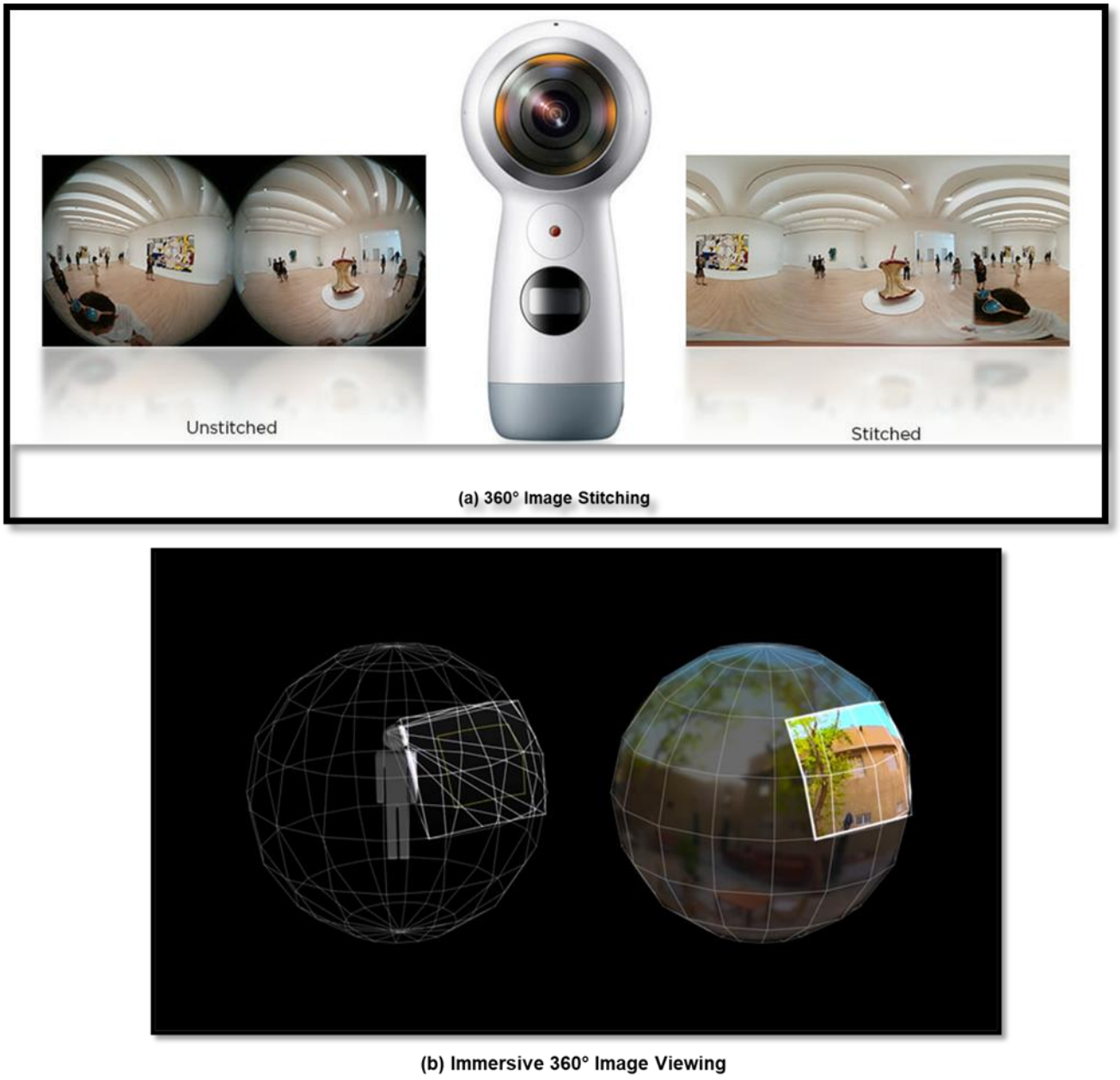


Figure 11. (a) An example showing how two 180° fisheye capture are stitched together in an equirectangular projection (Kuula Team, 2017). (b) An example of 360° VR is rendered and experienced inside HMDs (Insta360, 2023).

When compared to model-based VR, 360°-VR media provide a simpler workflow for development and a higher level of visual fidelity enhancing the psychological response from the users (Higuera-Trujillo et al., 2017). However, within game engines these assets are not processed as 3D objects, but as still images or videos. Thus, the only way users can interact with them is using 3-DoF exploration. Then, most 360°-VR experiences allow users to interact with these environments only using virtual buttons (i.e., hotspots) placed around the scene. As an example, for simulating teleportation usually these objects contain links to other 360° environments or to different POVs of the same scene. Lastly, 360°-VR scenarios are not as customizable as model-based environments, as their range of use depends only on what was recorded during development process. Thus, it is fundamental to program the recordings as precisely as possible.

2. Virtual Reality Technology in Neuropsychological Testing: A Systematic Review

2.1 Why VR for Neuropsychological Tests? Benefits and Challenges

Neuropsychology aims to describe and explain the relationships between the brain and behaviour, considering both normal functioning and pathological dysfunctions related to brain damage (VandenBos, 2015). The main method to investigate the presence and the extent of these impairments is called neuropsychological assessment, which can be defined as a "normatively informed application of performance-based assessments of various cognitive skills" (Harvey, 2012). This is typically done by assessing cognitive functioning with tests requiring the execution of behavioural tasks which are known to be related to some functions of the cognitive system (e.g., the memorization of a short story to measure memory).

One of the most distinctive traits of neuropsychology is being transdisciplinary (Casaletto and Heaton, 2017): as an example, its past dialogue with computer science and technology was fruitful both for the theoretical (e.g., connectionism, neural networks) and methodological (e.g., computerized paradigms) development of the discipline. The interchange is still meaningful today. For instance, VR technology recently emerged as an innovative approach that is potentially useful to improve the quality of neuropsychological tests. However, "paper-and-pencil" tests remain the gold standard for both neuropsychological research and clinical assessment (Parsons, 2015; Rizzo and Koenig, 2017).

This systematic review aims to examine the current state of VR technology in neuropsychological testing, including a discussion on the reasons for its introduction as a potential assessment tool and a brief history of its use in neuropsychological assessment. The review will present results and a discussion on the current state of the art, grouping the tools according to the cognitive domains they assess. The final section will offer a forward-looking perspective on potential future developments in the field.

In the scientific framework of behavioural studies, VR technology has been described as "an advanced form of human-computer interface that allows the user to interact with and become immersed in a computer-generated environment" (Ventura et al., 2018). This definition highlights the importance of immersion (i.e., the capability of a VR system to support the illusion of being present in the digital world), which is considered a technological feature provided by the interaction system (see Chapter 1). The extent to which immersion is supported by the VR setup defines its capability to simulate realistic experiences, which is one of the most important benefits that VR can bring to neuropsychological assessment. Chaytor and Schimtter-Edgecombe (2003) suggested that classical (i.e., paper-and-pencil) neuropsychological tests show only a moderate level of ecological validity, which refers to how much the information derived from the tasks performed in the laboratory setting can be linked to a person's real-life functioning (Tupper and Cicerone, 1990). VR technology allows the administration of behavioural tasks which can closely resemble their real-life counterparts. This is useful for observing user's ecological behaviour in a controlled and safe situation which can be otherwise difficult to organize due to logistic or safety reasons (Schultheis and Rizzo, 2002). Then, when compared to traditional techniques, VR allows the collection of more ecological measures of cognitive functioning, which can result in better prognostic indices, keeping at the same time a high level of experimental control (Parsons, 2016). In fact, VR experiences are designed using game engines software which enable full control of each virtual variable, both environmental (e.g., scene's objects) and user-related (e.g., reaction times), providing standardized stimuli and an error-free computation of the scores. This software also allows the creation of multiple versions (i.e., scenes) of the same task, thus potentially increasing the variability of the stimulation and its sensitivity in detecting changes over time (Serino et al., 2022). Moreover, VR hardware can provide real-time multisensory stimulation (e.g., visual, auditory, or tactile), which can assess cognitive functioning at different levels. Additionally, many of the currently available VR devices include different sensors allowing the collection of measures that are impossible to collect using classical assessment tools. For instance, some HMDs integrate an eye-tracker sensor that can measure the visual exploration

pattern (see Chapter 5). This possibility was exploited for detecting signs of sub-types of spatial neglect (Hougaard et al., 2021). Another fundamental source of diagnostic data relies on the motion-tracking capabilities provided by VR hardware. Miller and colleagues (2020) showed how the collection and elaboration of motion-related measures (e.g., head orientation or hand position) have the potential to identify "motion signatures" which are related to personal identity, medical condition or mental state. This possibility was indeed exploited to detect possible digital biomarkers related to specific signs of cognitive impairment (Cavedoni et al., 2020).

Finally, VR-related immersion has the potential to motivate the users and enhance their engagement during the assessment procedure, improving the reliability and validity of the evaluation (Bell et al., 2020). This can also be done by introducing elements coming from videogames, an approach called "gamification". In this case, the user's entertainment is not the aim of the experience, but it is a means which can prevent dropouts and support compliance with the neuropsychological assessment (Cabinio et al., 2020).

Although VR seems to provide clear methodological benefits to neuropsychological assessment, there are still some challenges to face. First, the exposition to VR can cause some negative effects, such as cybersickness. The most common symptoms include eye fatigue, sweating, headache, dizziness, and nausea. It seems to be related to an occurring mismatch between the afferent visual and vestibular signals (Gallagher and Ferrè, 2018), for example when our virtual avatar is moving forward while we are stationary in the real world. Even if the quality of VR hardware is improving and the best practices for designing VR environments are emerging (LaRocco, 2020), the susceptibility to cybersickness still relies on various human factors that cannot be completely controlled (Chang et al., 2020). Thus, this remains a risk that needs to be considered even more carefully in a context such as neuropsychological assessment, where often end-users are patients characterized by many frailties. To this end, despite VR systems have shown good usability and acceptability among elderly people (Tuena et al., 2020), their application may not be suitable for every assessment situation or every

user. As an example, the evaluation of patients with motor deficits or severe cognitive impairments still relies on the administration simplicity, which characterizes most paper-and-pencil tests that cannot be matched by VR systems. The feasibility of this transition must take into account the design constraints which are specific to the implementation of VR experiences for elderly people (Ijaz et al. 2022). Moreover, providing a high level of realism is very difficult and can require a great effort, both in terms of time and human resources (e.g., a team of specialized developers). Then, the three VR illusions are often only partially supported, leaving the users in the so-called "uncanny valley": that is, they start to feel sensations of eerie and uneasiness when interacting with the virtual world, specifically when the objects of the interaction are human-like virtual characters (Stein and Ohler, 2017). The level of realism provided is related to another important point: that is, the final monetary cost of these novel tools. Development costs surely impact the price of the final product: this is true both for new paper-and-pencil tests and VR tests, but for the latter, they might be larger. The inherent complexity of VR-based assessment tools may require more extended training for clinicians when compared to classical tools, increasing the final costs. Moreover, the cost of the whole VR package (i.e., devices and related software) must be considered. It is worth noting that commercial-level VR headset costs are rapidly and significantly decreasing and are now more affordable, with a range from around 300€ up to 4000€. The costs for the hardware are una tantum and thus can be absorbed by the usage of the instrument.

2.2 A Short History of VR in Neuropsychological Assessment

The first comprehensive contribution linking VR to neuropsychological tests is dated 2002. Schultheis and Rizzo (2002) discussed the main advantages of implementing this technology in the neuropsychology field, calling for a further effort to understand better the possible benefits of VR in the study of cognitive functions. Since then, authors have mainly focused on building prototypes of neuropsychological tests using VR and studying the feasibility of their application.

About 10 years later Bilder (2011), in its call for a paradigm shift in neuropsychology, proposed a "three-waves" review of the theoretical formulations that have occurred in the discipline, which was then revised in a more technology-oriented fashion by Parsons in 2016. Following Bilder's proposal, neuropsychology's methodological and technological advancements are divided into three subsequent temporal phases, each characterized by a specific assessment technique. The "classic" paper-and-pencil tests (i.e., those tests that do not require any technological equipment) belong to 1.0 neuropsychology. Computer-assisted tests characterize 2.0 neuropsychology. The 3.0 wave is now emerging with the development of more sophisticated tests based on simulation technologies and advanced data collection and processing (e.g., machine learning).

Novel VR tests belong to the 3.0 wave and were developed following two different criteria: a) adapting established assessment techniques in VR (e.g., a cancellation task, see Knobel et al., 2020); or b) creating virtual variants of real-life tasks (e.g., a virtual shopping task, see Ouellet et al., 2018). This dualism can be related to the scientific debate opposing emerging function-led approaches to traditional construct-driven approaches (Burgess et al., 2006). Briefly, tests developed with a construct-driven approach include tasks aimed to assess cognitive "constructs" (e.g., memory) without an explicit interest in predicting its declination as a functional behaviour, thus limiting the generalizability of the result obtained during the assessment to the real-life experience. On the other hand, function-led tests are developed using tasks which are representative of real-life-related functions (e.g., remembering the items to buy at a shop as a memory measure), giving the opportunity to observe the execution of complex behaviour as it would be performed in the everyday context. Thanks to the possibility of simulating everyday tasks, VR technology is functional to develop novel function-led tests. On the contrary, recreating construct-driven tests in VR can result in a mere superimposition of stimuli on the 3D VE, making these tests "little more than adaptations of outmoded conceptual and experimental frameworks", failing to represent the actual functioning capacities of the individuals (Parsons, 2015).

Over the past 20 years, there has been an increase in the number of studies on the use of VR in neuropsychological testing, with a particularly significant increase in studies on immersive VR since 2016 (Fig. 13). From 2016 to 2021, 134 studies that met the inclusion criteria were identified, while 154 studies were identified in the preceding 15-year period (2000-2015).

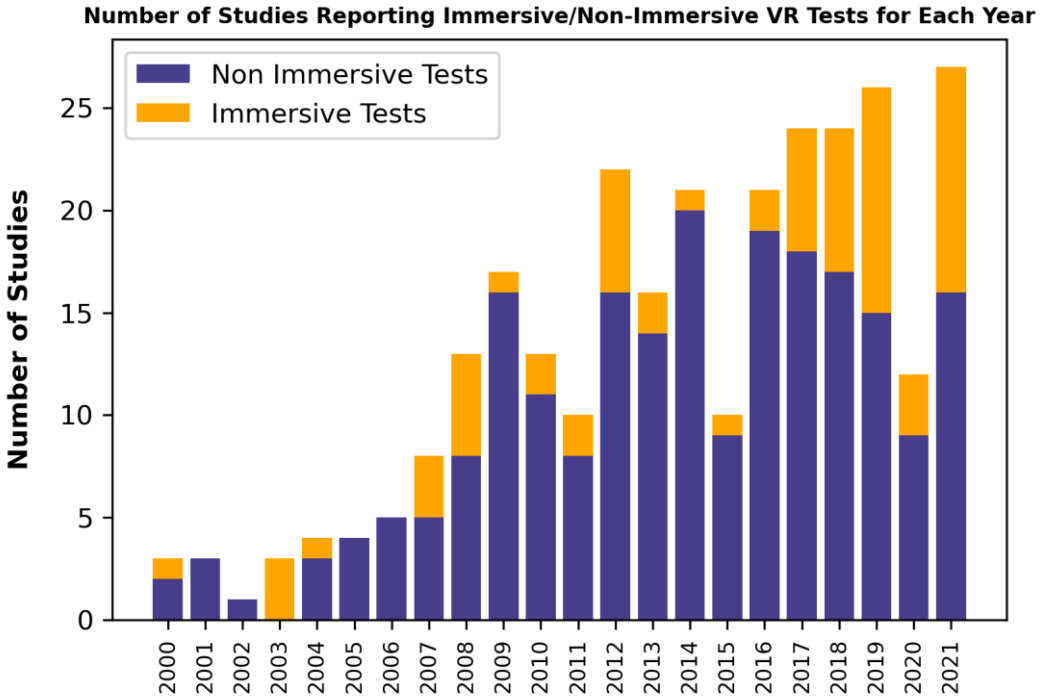


Figure 13. The plot shows the number of articles reporting immersive and non-immersive VR tests for each year. On the x-axis is reported the year of publication. The y-axis indicates the number of studies published. The purple bars represent the number of studies describing non-immersive tests, while the yellow ones represent the number of studies describing immersive tests. Image created with Spyder.

This rapid growth could be explained by the widespread availability of more consumer-oriented HMDs. Facebook and HTC/Valve launched Oculus Rift and Vive, respectively, giving the researchers the possibility of using relatively low-cost immersive solutions to improve the quality of VR protocols with more sophisticated settings. In parallel, 360°-VR raised the attention of the researchers due to the minor cost of the hardware, the visual fidelity of the environments and the reduced technical skills required (Serino and Repetto, 2018).

2.3 Methods

2.3.1 Search Strategy

This study was pre-registered in PROSPERO (ID: CRD42021257744). The literature review was conducted following the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA 2020) guidelines (Page et al., 2021). On 15/11/2021 searches were conducted in three databases, PubMed, Web of Science and PsychInfo, using the following search terms combinations: ("virtual reality" OR "VR") AND ("neuropsychological assessment" OR "neuropsychological test*"); ("virtual reality" OR "VR" OR "Virtual environment") AND ("neuropsychological assessment" OR "neuropsychological test*" OR "cognitive test*"). The presence of possible duplicates in the resulting collection was checked first using the Mendeley software automatic algorithm and then with the R package "Revtools" (Westgate, 2019), screening for identical titles and DOIs.

2.3.2 Study Selection

Articles published in international, peer-reviewed journals in English from 2000 to 2021 were considered for inclusion in the study. Studies that (1) were not based on VR protocols, (2) were based on VR but did not concern neuropsychological tests, (3) described only rehabilitation or training protocols, (4) concerned reviews or metanalysis and (5) reported case-studies were excluded.

Stricter inclusion criteria were applied to identify articles that reported normative data for VR-based tests. The actual integration of the test in the clinical setting indeed relies on the availability of the normative data coming from the administration of the test to a large sample of healthy participants so that a reliable distribution of healthy behavior can be outlined, and a deviation from it can be identified (Crawford and Howell, 1998).

2.3.3 Data Extraction

Each eligible full-text article was analyzed to retrieve the following data: name of the first author, year of publication, size of the included samples, type of VR (immersive vs non-immersive), cognitive functions tested, the clinical condition studied (if any), name of the tool, test availability (on request, online or commercial) and the eventual presence of normative data related to the test described in the article. When available, evidence of test validity (e.g., convergent correlation with similar tasks) has been reported as well. Cognitive functions tested were described based on the indications provided by the original authors, but this information should be interpreted with caution as there is no consensus on the specific standard tests that assess particular cognitive functions (Cox et al., 2013), and even less for VR-based tests because they often include complex and dynamic tasks. Neuropsychological tests do not assess a single function; instead, they elicit a set of interrelated processes which can be linked to a predominant function (Tosi et al., 2020).

VR tests were then classified as "immersive" or "non-immersive" according to whether they used an HMD or not (Rizzo and Koenig, 2017).

2.4 Results

The PRISMA flow diagram sums up the literature review process (Fig 14). The research string resulted in 258 studies being included in the review and in 3 studies reporting normative data.

2.4.1 VR Tests Providing Normative Data

Despite a large number of papers, only three instruments are accompanied by studies reporting adequate psychometric proprieties (i.e., validity, reliability, sensitivity measures) and normative data, claiming to be ready for use in everyday clinical assessment: the Nesplora AULA (Iriarte et al., 2016) and the Aquarium (Climent et al., 2021), both assessing attention (see section 4.4) and the Systemic Lisbon Battery (Gamito et al., 2016a; 2016b) assessing memory and attention (Table 1). Notably, the Systemic Lisbon Battery reports a relatively small normative sample, and its clinical use may benefit

from additional validation studies. The two Nesplora instruments, instead, provide data on an adequate sample size for clinical use.

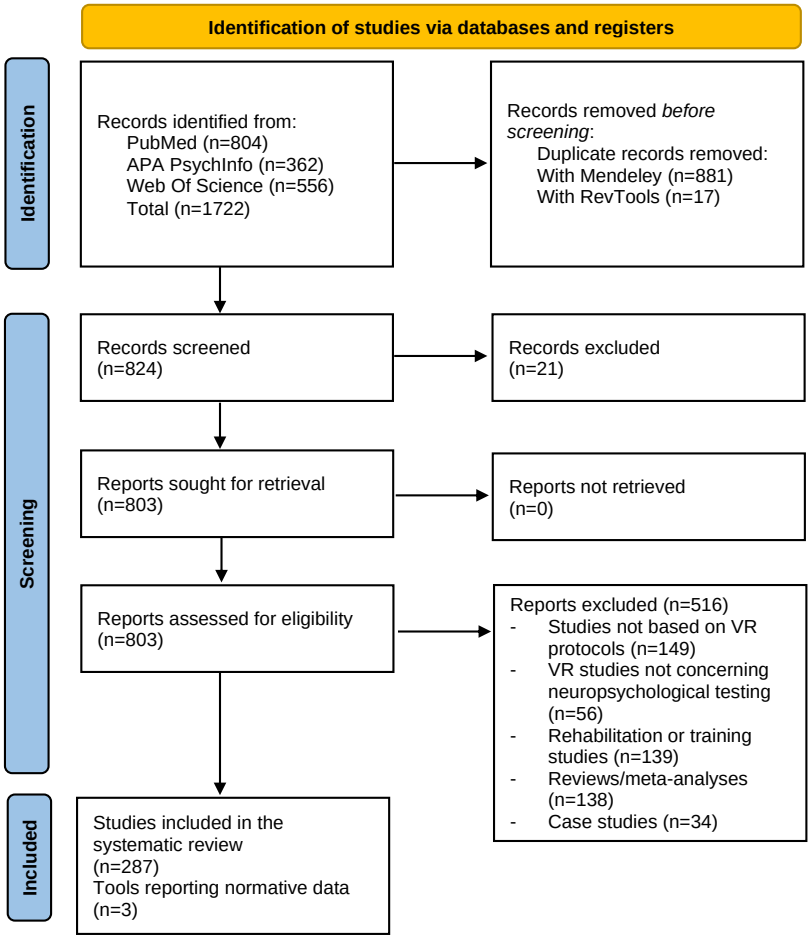


Figure 14. The PRISMA diagram resuming the flowchart of the systematic review.

First Author	Publication Year	Sample Size	VR Type	Cognitive Domain	Tool Name	Availability
Climent	2021	1469	Immersive	-Attention -Working Memory	Nesplora Aquarium	Commercial
Gamito	2016	243	Non-Immersive	-Attention -Memory	Systemic Lisbon Battery	Online
Iriarte	2016	1272	Immersive	-Attention	Nesplora Aula	Commercial

Table 1. The table provides a summary of studies that reported normative data for VR-based neuropsychological tools. It includes the name of the first author, the year of publication, the sample size, the type of VR technology (immersive or non-immersive), the cognitive domains measured by the test, the name of the tool, and the availability of the tool (on request, online, or commercial).

2.5 Discussion

2.5.1 Cognitive Domains

To provide an overview of the current state of VR neuropsychological testing, the review was extended to include tests without normative data. VR tests have been developed to study various cognitive domains (e.g., spatial navigation, memory, attention), and some areas appear to be particularly suitable for exploration with these tools (Fig. 15). Examples of VR tasks for each cognitive domain are provided below. This review focuses on cognitive domains traditionally related to assessment, while other aspects of cognition (e.g., body representations) studied with VR can be found in Serino et al. (2022).

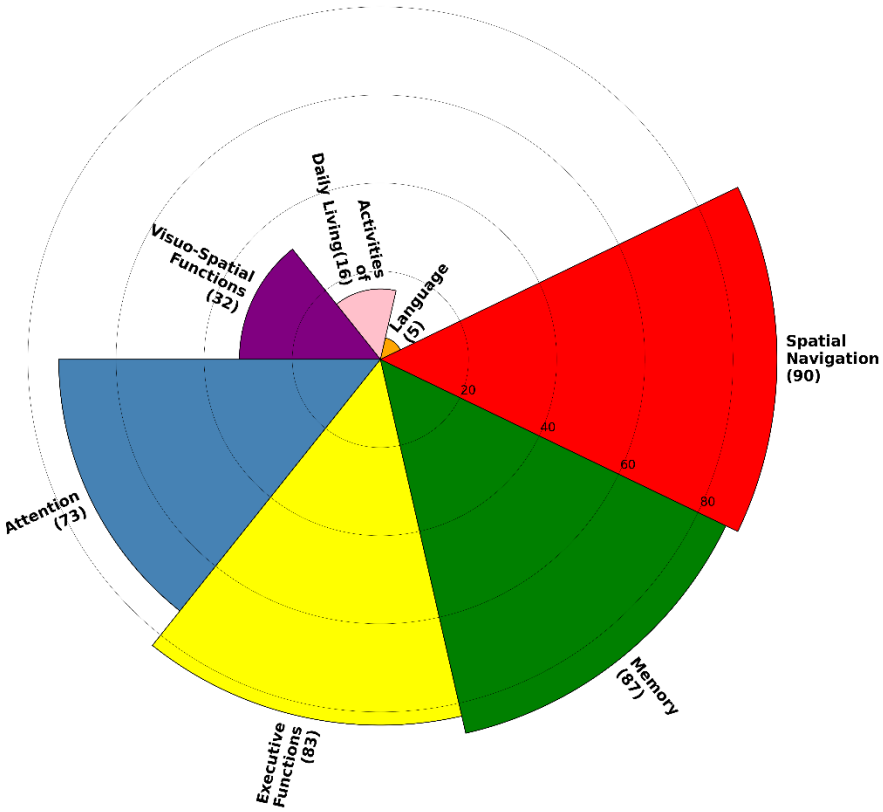


Figure 15. The polar chart shows the number of studies per cognitive domain. The numbers in parenthesis represent the absolute frequency of tests considering each function. Notably, the total frequency size (n=386) is larger than the total number of studies included in the 2000-2021 timespan (n=287). That is because many VR tasks are associated with the measurement of more than a single cognitive domain. Image created with Spyder.

2.5.2 Spatial Navigation

Spatial navigation can be defined as the ability to remember and navigate environments (Stuchlik et al., 2014). The construct is based on the studies of rodents using mazes (Morris et al., 1982). VR emerged as a powerful tool to develop complex environments which can be easily manipulated to match the experimental needs. For example, Machado and colleagues (2019) presented the VRmaze, a powerful software that includes a set of 3D non-immersive environments resembling the three most widely used experimental mazes: the reverse T-maze (Packard and McGaugh, 1996), the Morris Water Maze (Morris, 1982) and the eight-radial arm maze (Olton and Samuelson, 1976). These three environments are fully customizable, allowing full control of the maze building. This example shows how VR can bring complex environments into the laboratory. This has been used in healthy people (Martelli et al., 2019), but it can also assess spatial navigation in cognitively impaired patients in a safe and controlled way. Navigation impairments are a key symptom of many clinical conditions, such as traumatic brain injuries (Livingstone and Skelton, 2007). VR environments proved to be sensitive in detecting those deficits, discriminating people presenting with mild cognitive impairment from healthy elders (Weniger et al., 2011), or differentiating patients with Parkinson's disease from healthy controls (Schneider et al., 2017).

2.5.3 Executive Functions

Executive functions are a set of cognitive processes responsible for goal-directed behaviors (Barkley, 2012). It is often considered an umbrella term that includes planning, problem-solving, inhibition, multitasking and cognitive flexibility. Given that complexity, the neuropsychological assessment of executive functioning with traditional tests faces many difficulties, often related to the low predictive value of the measures (Chan et al., 2008). VR tasks can match the level of complexity required for engaging the highest possible number of executive components simulating complex real-life situations, leading to a more adherent measure of patient capabilities. For example, Raspelli and colleagues (2011) developed a Virtual version of the Multiple Errands Test (MET) (Shallice and

Burgess, 1991). The vMET can be considered a good representative of VR tests for executive functions. The original, non-virtual, Multiple Errands Test (MET) was developed to assess executive functioning in real-life settings (e.g., the hospital). Patients are asked to perform simple daily life tasks (e.g., buying a loaf of bread), with the experimenter that follows the situation and records the performance. The original MET has shown a consistent level of ecological validity; however, it is extremely time-consuming and complex to set up. On the contrary, its virtual counterpart overcomes these difficulties by mimicking the original tasks through a PC screen and a joypad. Scores computation includes a comprehensive set of information (e.g., process speed, error). Preliminary results showed significant correlations (r = from 0.70 to 0.80) between the vMET and measures of executive functioning included in the Test of Attentional Performance (TEA) (Zimmermam and Fimm, 1992), but not with other established measures of executive functioning such as the Iowa Gambling Task (IGT) (Bechara et al., 1994) and the Stroop Test (Stroop, 1935). Additionally, the vMET sensitivity in detecting executive impairments was reported across different clinical conditions, such as obsessive-compulsive disorder (La Paglia et al., 2012) and Parkinson's disease (Cipresso et al., 2014).

Davison and colleagues (2018) developed three immersive virtual real-life tasks (i.e., parking a car, arranging and locating items in a chemistry lab) requiring participants to ignore irrelevant information and focus on the specific goal. Results showed that TMT-A scores correlated significantly with the parking tasks ($\tau = 0.34$, $\tau = 0.38$ and $\tau = -0.49$), the arranging tasks ($\tau = 0.37$, $\tau = 0.25$ and $\tau = 0.3$) and the item location task ($\tau = -0.38$). TMT-B scores were significantly related to the number of levels completed of the parking simulator task ($\tau = -0.38$). Finally, Stroop Test scores were significantly associated to the performance on the parking simulator tasks ($\tau = -0.32$, $\tau = 0.43$), on the seating arrangement tasks ($\tau = -0.39$, $\tau = -0.33$ and $\tau = 0.33$ and $\tau = -0.37$) and on the item location task ($\tau = -0.37$).

In addition, Serino and colleagues (2017) showed the feasibility of using 360°-VR for the assessment of executive functions by developing the Picture Interpretation Test-360° (PIT-360°). The PIT-360° is a virtual adaptation of the standard PIT described by Rosci and colleagues (2005). In this test, the participants are asked to interpret a virtual scene resembling the classic test of Favretto's painting called "Il Sorcio". The scene was presented around them through an HMD. The PIT-360° was able to differentiate patients presenting with Parkinson's disease from control healthy subjects. Notably, the PIT-360° correlated with paper-and-pencil measures of executive functioning such as the TMT-A ($r = 0.50$) and the phonemic fluency task (Carlesimo et al., 1996) ($r = -0.57$).

2.5.4 Attention

Attention can generally be defined as "the focusing of our inner resources and state of consciousness" (Cohen, 2014). Attentional processing has been among the first functions to take advantage of technological support by adopting computer-based tasks like the measure of simple reaction times.

These methods are still relevant for basic research, but they are limited to generalizing to the real-world, (Parsons and Rizzo, 2008). VR can be used to mimic different real-life environments characterized by highly ecological stimuli or distractors (e.g., a crowded street). An example of how VR can be used for the assessment of attention is given by the Art Gallery Test (AGT), developed by Gamito and colleagues (2017). This is a non-immersive VR tool based on the comparative visual search paradigm (Pomplum et al., 2001). The task requires the participants to compare shapes presented in the two halves of a visual display, and it is considered a measure of selective attention processes. The AGT adapts and extends this paradigm by adding an ecological VE simulating an art gallery. Three different tasks are performed in the virtual gallery using the paintings as stimuli: a spot the differences task (AGT-A), a completion task (AGT-B) and a visual search task (AGT-C). Results showed significant correlations between the AGT-A and the interference index of the Color Trails Test (CTT) (D'Elia et al., 1996) with $r = 0.45$, and between the AGT-C and the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005) with $r = -0.42$.

The engaging nature of VR experiences is important for enhancing the ecological value of the obtained measures but could also be important for strengthening the compliance of the tested individual. This is particularly relevant when testing children populations and was thus exploited by authors interested in studying or assessing attention deficit hyperactivity disorder (ADHD). The proficient tradition of VR test for ADHD assessment was started by Parsons and colleagues (2007), who developed an immersive virtual scenario called Virtual Classroom, which included ecologically relevant adaptations of two tests, the Continuous Performance Task (CPT) (Riccio et al., 2002) and the Boston Naming Test (BNT) (Kaplan et al., 1983). The VE resembled a classroom where a virtual teacher gave the participants instructions for each specific task. The scores in the virtual classroom correlated with prototypical ADHD assessment measures such as the SWAN Behavior Checklist (Swanson et al., 2001) and Conners' CPT-II (Conners, 2000) with a moderate-to-large effect size (r ranges= 0.39 and 0.79). Moreover, the virtual class was able to distinguish ADHD from non-ADHD participants (Parsons et al., 2007). The virtual class is a first but not a unicum (Bioulac et al., 2012; Seo et al., 2019). The immersive Nesplora AULA test (Iriarte et al., 2016) can be considered one of the most complete tests in VR. It is based on the Conners' CPT paradigm, set in a virtual classroom environment that requires the participants to respond (1) when any letter except the letter "X" or (2) when the letter "X" appears on the blackboard. A complete psychometric approach has been used to validate the test, reporting evidence for its validity in the assessment of attention in children from 6 to 16 years old (Díaz-Orueta et al., 2014; Areces et al., 2018a, 2018b; Zulueta et al., 2019). The test disposes of reliable normative values obtained on large sample sizes for the administration in the Spanish population (Iriarte et al., 2016), which makes it one of the few VR-based tests ready to be used in everyday practice (at least in Spain). Notably, the same group recently provided the normative values for another Nesplora test: the Aquarium, which can be used to assess attention in people older than 16 years old (Climent et al., 2021). This test too is based on CPT paradigms, but it is performed in a virtual aquarium and the target/non-target stimuli are based on fishes' names and appearances.

2.5.5 Memory

Memory includes all those processes which allow us to encode, store and retrieve information (Squire, 2009). Recent classification divides this construct into sensory, short-term, and long-term (Camina and Güell, 2017). The dynamic properties of VR can be fruitfully exploited to study the composite and complex nature of memory processes. For example, innovative VR procedures can be used to easily manipulate the perspective in which participants see virtual scenes through an HMD. This possibility can be very useful for the study of memory processes, as proved by Bergouignan et al. (2014), who showed how encoding information using a first-person (i.e., egocentric) or a third person (i.e., allocentric) perspective affects episodic memory. This is true also for a higher level of mnemonic processing, namely the autobiographic memory, which also encompasses a larger set of cognitive operations (e.g., self-reflection, emotional engagement) (Schöne et al., 2019) that can be elicited by the immersion provided by VR. For example, Kisker and colleagues (2021) showed that immersive VR promotes autobiographical retrieval mechanisms, consequent to an encoding process closer to real-life events.

Memory seems a promising field of application where virtual mazes have already been used both in healthy (Tascòn et al., 2017; Tascòn et al., 2019) and clinical (Cànovas et al., 2009; Spieker et al., 2012; Ritchie et al., 2018; Wilkins et al., 2013) populations.

A pioneering study in the field was conducted by Matheis and colleagues (2007), who developed the VR Office, a task that required the participants to perform a list-learning memory test within a simulated and immersive office environment. The authors showed how the VR Office test distinguished traumatic brain-injured patients from controls; additionally, the test correlated ($r = 0.70$) with the California Verbal Learning Test (CVLT) (Delis et al., 1987), providing convergent validity indices.

Ten years later, despite the promising results, only a few VR tests are focused on memory assessment. The Virtual Shop (Canty et al., 2014; Parsons and Barnett., 2017) is a pilot VR test for an ecological

evaluation of everyday memory (Corriveau Lecavalier et al., 2018; Ouellet et al., 2018). The task is administered in an immersive VE where the participant is first asked to memorize a list of items to buy and then find them in the simulated, realistic virtual shop (e.g., presenting distractors like a conversation between two people or similar products). The virtual shop is ageing-sensitive and correlates with a traditional free recall word list test (Belleville et al., 2003), both for immediate (r = 0.57 for young and 0.28 for elders) and delayed recall (r = 0.46 for young and 0.30 for elders). In the same year, Serino and Repetto (2018) discussed how immersive VR based on 360°-VR could be used as a novel tool for memory assessment. Negro Cousa and colleagues (2019) used the same technology for the assessment of memory functions. Results cautiously indicated that 360°-VR environments enhance memory encoding, opening a new window for VR setup in neuropsychological testing.

2.5.6 Visuospatial Functions

Visuospatial functions are those processes responsible for stimulus identification and localization in the space, allowing the understanding of visual representations and their spatial relations (Très and Brucki, 2014). The visuospatial functions can be divided into different components, commonly including visual perception, construction, and visual memory (Lezak, 2012). The study and assessment of these components traditionally included paper-and-pencil tests or 2D computerized tasks requiring mentally manipulating object configurations (Vandenberg and Kuse, 1978), and drawing or copying figures (Rey, 1941). VR implementation is based on common paradigms that include navigational tasks performed in highly ecological environments and situations. For example, Buxbaum et al. (2008) developed a Virtual Wheelchair Navigation Test (VWNT). It is a non-immersive VR setup where participants are required to name objects appearing on both sides of the display to avoid the virtual collision. The VWNT includes four conditions differentiated by the complexity of the tasks (e.g., more objects to name or less guided navigation). Results showed that the VR task could differentiate patients from controls and strongly correlated (r = 0.85) with a real wheelchair navigation task. These results were replicated in a similar study (Buxbaum et al., 2012)

adopting a slightly different task (Dawson et al., 2008). In the same framework, Navarro and colleagues (2009) presented the Ecotrain-Cognitive, a non-immersive task where participants needed to navigate safely from different locations placed in a simulated city and avoid accidents with the passing cars. The VR task correlated to the Behavioural Inattention Test (Wilson et al., 1987) (r= -0.50) and the time to complete the CTT (r = 0.70 for part A and r = 0.80 for part B). Similarly, Kim et al. (2010) implemented a measure of deviation angle on a virtual driving task which correlated with the performance of neglect patients on a line bisection task (r = 0.63).

A parallel line was developed by adapting classical tests for neglect (e.g., cancellation tasks). The adaptation used complex 3D stimuli and environments, allowing for a more advanced and complex investigation of neglect-related impairments (Fordell et al., 2011; Pallavicini et al., 2015; Knobel et al., 2020).

2.5.7 Activities of Daily Living (ADL)

Activities of daily living (ADL) include two groups of skills that are crucial for independent self-caring: the Basic ADL (BADLs), like eating, bathing, or dressing and the Instrumental ADL (IADLs), which includes more complex activities like managing finances or medications and doing grocery (Katz et al., 1963). In neuropsychological evaluation, the ADL assessment is crucial for understanding if the person can take care of her/himself and face the challenges of everyday life. Information about ADL is generally obtained via indirect sources, such as interviews or questionnaires administered to individuals and their caregivers (Isik et al., 2020). VR makes it possible to simulate realistic everyday life environments and related challenges, directly measuring ADL and outperforming the clinical predictive validity of traditional assessments (Tarnanas et al., 2013). VEs simulating IADL assessment have been developed for different levels of complexity, like kitchens (Allain et al., 2014; Besnard et al., 2016; Richard et al., 2010; Vallejo et al., 2017; Zhang et al., 2003), house rooms (Albani et al., 2002; Ruse et al., 2014), or external scenarios like streets and

cities (Atkins et al., 2018; Seo et al., 2017), however, the convergent validity of these measures still need to be demonstrated.

2.5.8 Language

Classical neuropsychological paradigms for studying and assessing language are based on tasks investigating written and oral comprehension, production, and repetition. In respect of other cognitive domains, language functions are poorly studied in VR: in fact, many of the existing VR tests do not include specific tasks or scores for the language evaluation (Oliveira et al., 2016; Wall et al., 2017) and the language domain is almost untreated by the current literature of VR assessment.

2.6 Conclusion

This systematic review aimed to provide an up-to-date description of the current state of VR neuropsychological testing and to identify strengths and weaknesses. In the last twenty years, many contributions supported VR-based neuropsychological tests' feasibility, and the amount of evidence is proliferating. Technological advancements are emerging and spreading thanks to the lowering costs and greater accessibility of VR hardware and software.

VR offers the opportunity to reformulate the neuropsychologist's testing tools, improving methodologies and surpassing old testing problems, such as the low ecological validity of the assessment procedures. Notably, still today, some VR tests are conceived as enhanced adaptations of classical paper-and-pencil tests. While this approach can be relevant for improving stimuli presentation and response coding, it does not take full advantage of what VR can offer (Parsons, 2016). It may indeed limit the potential of this new technology, which may offer a more radical exploitation potential. Thus, VR should not be used to build tests that aim to substitute the standard ones as mere high-tech fancy new tools. On the contrary, VR tests should provide alternative evaluation tools complementary to the standard ones. For example, VR tests have the potential to improve the ecological validity of testing by simulating realistic experiences. But they could also be

used to provide novel information based on computerized data, like virtual space exploration strategies (e.g., heatmaps, exploration trails). They can also stimulate different (or multiple) sensory modalities with precision and control.

VR tests may also support a new perspective of the assessment, searching for what is preserved and not for what is impaired in a patient, suggesting, by means of real-life simulations, where an individual is independent and where he/she is not. This approach, on the other hand, may lose the specificity of the impaired function. For this reason, it could be interesting to parallel the approaches where both ecological and formal testing is administered.

Importantly, almost every VR neuropsychological test remained bound to research applications. The development of these tools often stops at a preliminary stage, with the majority of studies reporting evidence for their feasibility. Psychometric validity and sensitivity are rare, while only a few recent tools are also accompanied by normative studies reporting adequate sample sizes (Iriarte et al., 2016; Climent et al., 2021). Normative values are essential for clinical use, but they require data collection from hundreds or thousands of people, requiring much effort and time from the researcher's perspective. Moreover, many of the considered articles provided validity of novel VR-tool showing their correlation with traditional neuropsychological tests (i.e., convergent validity). This approach may be biased if used to demonstrate the adequacy of VR tests in detecting real-life impairments. In fact, it assumes that standard tests are the gold standard for this aim, while it was shown that their ecological validity is limited (Chaytor and Schmitter-Edgecombe, 2003). Thus, if a VR test correlates weakly with a traditional test, it does not necessarily imply that the test is not satisfactory. Moreover, the discussion concerning the improvement of ecological validity is still mainly theoretical. Future studies should focus on testing and demonstrate how VR is better in estimating-real life activities, both in terms of veridicality (i.e., the degree to which test scores correlate with measures of real-world functioning) and verisimilitude (i.e., the degree to which tasks performed during testing resemble those performed in daily life) (Parsons, 2016).

Additionally, the implementation of VR technology as an assessment tool is not risk-free and may not be suitable for every assessment situation (e.g., severely impaired patients). Researchers and developers must carefully consider the challenges of using VR with people who have cognitive impairments, including usability issues and potential side effects such as cybersickness. On the technological side, the growing availability of consumer-oriented HMDs allowed the researchers to afford cheap and ready-to-use immersive solutions that can be used for developing highly ecological VR-neuropsychological tests.

In conclusion, a growing body of evidence shows the potential offered by the implementation of VR technology in developing innovative neuropsychological tests. The 3.0 tools are nowadays a feasible reality. It seems that they have reached a level of scientific maturity that can also extend their usefulness from the lab to the ambulatory, with the development of advanced and ecological assessment tests that can improve the quality of classical neuropsychological evaluations. However, a strong effort to obtain reliable psychometric evidence is very much needed in the next years to reach this goal.

3. Model-Based VR vs 360°-VR: A Comparison of Effects on Task Performance and User Experience

3.1 Theoretical Background

The previous Chapter showed how VR environments are currently used to develop more sophisticated tools for assessing cognitive functions. The benefits and challenges this novel technology bring within neuropsychology must be carefully weighted to find new ways to build increasingly functional and safe 3.0 neuropsychological tests. This also includes exploring the feasibility of adopting solutions which emerge from technological advances.

In 2017 Silvia Serino and colleagues were among the first to implement a new kind of VEs in neuropsychological assessment based on spheric image or video captures of real-world scenarios (i.e., 360°-VR, see Paragraph 1.3.2). The main advantages offered by this media concerned a simpler workflow for implementation and an improved level of visual realism (Serino and Repetto, 2018). The latter was immediately seen as an interesting feature for neuropsychology, as it could have a significant role in providing virtual test environments characterized by a higher level of ecological validity. However, there has been debate about whether spherical 360° environment should be considered part of VR technology, mainly due to its limited interactivity with the virtual world. Slater and Sanchez-Vives (2016) suggested that, at least for 3-DoF applications, 360°-VR can elicit sensorimotor contingencies comparable to those that characterize Model-Based VR environments. Thus, given its impossibility of supporting 6-DoF exploration, 360°-VR can be considered less immersive than Model-Based VR in absolute terms but can still be fully considered part of the VR spectrum.

Given its photorealistic nature, by definition, 360°-VR is theoretically considered superior to Model-Based VR in providing ecological stimuli for assessment, although this claim is yet to be evaluated empirically. In fact, in recent years, a growing number of studies successfully piloted 360°-VR in the field of neuropsychological assessment and showed preliminary evidence supporting its use as a tool

for measuring cognitive performance (Serino et al., 2017; Realdon et al., 2019; Pieri et al., 2021; Bruni et al., 2022). However, a study including a direct comparison between 360°-VR and Model-Based VR is still missing.

The aim of the present work is double: exploring the possible influence of VE types on task performance and user-experience, including perceived ecological validity. To do so, an immersive Target Selection (TS) Task was implemented using Model-Based VR and 360°-VR. The task is inspired by cancellation tasks used for testing attentional performance in clinical settings (Brucki and Nitrini, 2008) but revised more ecologically.

First, results concerning the performance differences detected in the two conditions are described, followed by an analysis of user-related experience (Experiment One). Then, a second study with the same structure and objectives but using a Free Visual Exploration Task (FVE) is presented (Experiment Two). Finally, an integrated view of both experiments is shown at the end of the Chapter.

3.2 Experiment One: Target Selection Task

3.2.1 Methods

3.2.1.1 Participants

In this between-subject study, 99 participants were recruited through advertisements within the University of Milano-Bicocca. They were randomly assigned to two groups, one for the 360°-VR condition and one for the Model-Based VR condition. The study was conducted in accordance with the Helsinki Declaration of 1975, as revised in 2013 and approved by the Ethics Committee of the Department of Psychology of Milano-Bicocca. The exclusion criteria for this study included impairments that could have made it unsafe or uncomfortable for participants to be exposed to VR, such as a severe internist, psychiatric, neurological impairments or having a previous history of motion sickness symptoms. The samples included in the statistical analysis consisted of a group of 50 participants (31 females, 18 males, 1 non-binary) for the 360°-VR condition and a group of 49

participants (32 females, 16 males, one non-binary) for the Model-Based VR condition. The mean age of participants within the 360°-VR group was 22.8 years (SD = 5, range = 19-52), and the mean years of education were 14.08 years (SD = 1.44, range = 13–16). The mean age of participants within the Model-Based VR group was 22.9 years (SD = 3.2, range = 19-35), and the mean years of education were 14.6 years (SD = 1.49, range = 13–16). All participants were provided with detailed information about the procedures, potential risks, and data usage, both verbally and in written form. After being fully informed, participants were asked to provide written consent to participate in the study.

3.2.2.2 Application Design and Task Administration

Two separate immersive applications were developed, one for each condition. These applications shared a common initial scene, which functioned as a familiarization phase. In this scene, the user starts in the middle of a basic virtual room and hears audio instructions concerning the familiarization task and the subsequent TS. To keep the instructions as standard as possible, the corresponding audio files were created using Balabolka, a text-to-speech software based on Microsoft Speech API. The familiarization task consists in searching the environment and selecting through a raycast three icons showing the numbers 1-2-3 (Fig. 16), which are positioned on the right, behind and on the left of the starting position, respectively. This procedure has been used previously in a similar application (Pieri et al., 2021) and allows the users to familiarize themselves with 3-DoF exploration and user interface interaction. Once the last number is selected, the test phase begins.

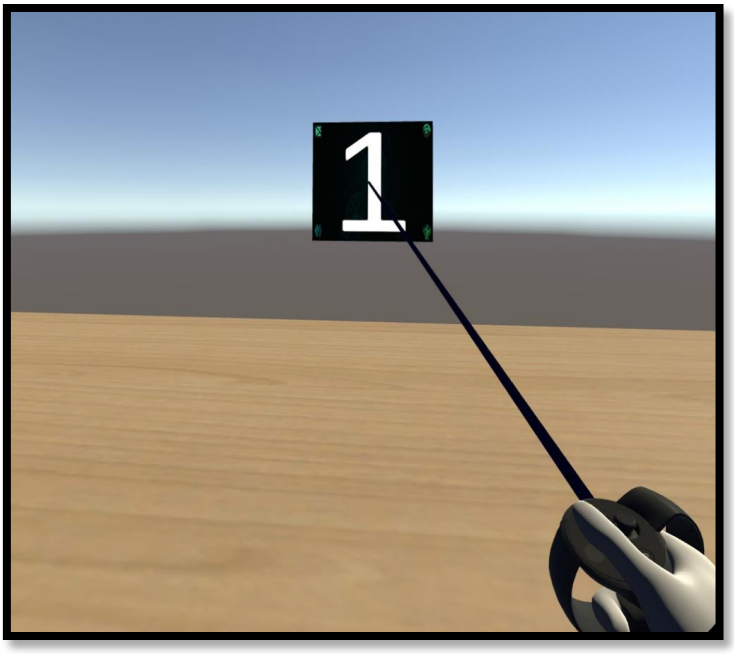


Figure 126. A screenshot showing the raycast interaction during the familiarization phase. The number "1" is positioned on the right of the starting position. The subsequent "2" and "3" are placed behind and on the left.

In both test conditions, participants are placed in the center of a living room with plastic cups and glasses scattered around. They are instructed to dispose of some of these items in the trash. To do this, they need to quickly find and count only the yellow cups among the 36 total stimuli, which include 11 yellow cups, 21 green cups, and 4 glasses.

The two applications were developed using Unity game engine with the Oculus Integration SDK Version 29.0, a package which includes fundamental prefabs such as models, animations, and interaction scripts, all specifically designed for working with the HMD chosen for delivering both applications, the Meta Quest 2. This headset integrates a screen with a resolution of 1832x1920 pixels per eye and supports both 3-DoF and 6-DoF exploration, either in stand-alone or as PC-tethered (see Paragraph 1.2.2). To ensure the best possible visual quality and smoothness during the experience (i.e., a stable framerate of 90 FPS at runtime), the target platform chosen for delivering the application included tethering the Quest 2 with a high-performance HP Omen Desktop PC with Intel i9 CPU, NVDIA RTX 3090 GPU and 64GB of RAM.

For the 360°-VR variant of the task, a spherical photo of the living room set was captured using an Insta360 One X camera. The resulting photo had a resolution of 5.7K (5760x2880 pixels) and was stitched in an equirectangular projection using Insta360Studio. This image was imported in Unity as a non-compressed texture, which was used to fill the internal face of a sphere object using a custom shader C# script, positioning the camera origin at the centre of this sphere (see Paragraph 1.3.2). On the other hand, the model-based environment was created using Unity by importing 3D models and arranging them to match the appearance of the original 360°-VR scenario closely. A visual comparison of the two environments is presented in Figure 17.



Figure 17. A comparison of the Target Selection Task between the two modalities. The upper image shows the 360°-VR variant of the task, while the lower image shows the Model-Based variant of the task.

3.2.2.3 Administered Questionnaires

Questionnaires administered to participants aimed to collect data concerning different dimensions related to user-experience.

The primary self-report measure used to gather information on the experience related to the exposition in the two different media was the ITC-Sense of Presence Inventory (ITC-SOPI), developed by Lessiter et al. (2001). This questionnaire includes 44 items which are worded such that a higher score on each dimension indicates a greater intensity related to the experience.

Participants are asked to rate their agreement with each statement on a 5-point Likert scale, with 1 indicating "strongly disagree" and 5 indicating "strongly agree." The ITC-SOPI questionnaire measures 4 dimensions of sense of presence in a VE:

• Spatial Presence: This dimension assesses the extent to which a person feels as if they are physically located within the VE. It evaluates factors such as the realism and believability of the VE, the participant's ability to orient themselves within the environment, and the extent to which the participant's physical body is perceived as being present in the VE.

• Engagement: This dimension measures the level of involvement and interest a person has in the VE. It assesses the extent to which the participant is absorbed in the VE, how much the participant feels part of the VE and how much the participant is able to focus on the VE.

• Ecological Validity: This dimension evaluates how realistic the participant perceives the VE. It examines factors such as the similarity between the VE and the real world, the consistency of physical laws within the VE, and the overall believability of the VE.

• Negative Effects: This dimension measures any negative aspects of the VE experience. It evaluates factors such as disorientation, discomfort, cybersickness, and any other negative effects that the participant may have experienced.

The ITC-SOPI was widely used in many studies evaluating users' responses to VR media, showing good psychometric proprieties (Vasconcelos-Raposo et al., 2018; Yan et al., 2020).

To control for individual differences in immersion tendencies that may affect the results of the study, the Immersivity Tendencies Questionnaire (ITQ) developed by Witmer and Singer in 1998 was also administered. The ITQ aims to assess an individual's tendency to become easily immersed or absorbed in different types of activities, such as VR experiences or video games, to ensure that any immersion experienced during the study is a result of the VE and not the participant's tendencies. The questionnaire includes 18 items rated on a 5-point Likert scale, with 1 indicating "strongly disagree" and 5 indicating "strongly agree".

Finally, the System Usability Scale (SUS, Brooke, 1986) was administered to test how participants evaluated the interaction between the two VR applications. The SUS consists of 10 items that are rated on a 5-point Likert scale, with 1 indicating "strongly disagree" and 5 indicating "strongly agree". The items are designed to measure the user's perception of ease of use, ease of learning, and perceived satisfaction with the system. The SUS produces a single score, known as the SUS score, which ranges from 0 to 100, with higher scores indicating higher perceived usability.

3.2.2.4 Experimental Procedure

Once the written consent was obtained, participants were invited to sit in a swivel chair and instructed to wear the Meta Quest 2 headset. The experimenter then launched the application on a desktop PC and remained in the room to collect the participant's responses and provide assistance if necessary. The user was guided through the VR experiences by audio instructions embedded within each scenario. After task completion, the participants were asked to relax for a few minutes before filling the questionnaires.

3.2.2.5 Statistical Analysis

A Bayesian approach was chosen for statistical analysis, as it has been shown to provide advantages over frequentist statistic in psychology (van De Schoot et al., 2014). Bayesian statistics is a method of analyzing data that updates the probability of a hypothesis based on added information. It assigns probabilities to hypotheses and adjusts them as new data is collected. This approach allows for the inclusion of prior knowledge and the quantification of uncertainty, providing a more comprehensive understanding of the data and hypothesis being studied (Wetzels et al., 2011; Kruschke, 2014; Drovandi et al., 2015). In Bayesian statistics, prior probability represents the initial belief about the value of a parameter before any data is collected. It is a probability distribution that represents the current state of knowledge about the parameter. The posterior probability, on the other hand, is the updated belief about the value of the parameter after data has been collected. It is calculated by combining the prior probability with the likelihood of the data. It represents the updated state of knowledge about the parameter considering both the preceding information and the new data.

In Bayesian statistics, it is possible to compare the evidence in favor of different hypotheses by computing the ratio of their posterior probabilities, known as the Bayes factor (BF). The BF provides a measure of the evidence in favor of one hypothesis over another. For what concern notation, BF_{10} defines the evidence in favor of the alternative hypothesis (H_1) over the null hypothesis (H_0). In contrast, BF_{01} defines the evidence in favor of the null hypothesis over the alternative hypothesis (Keysers et al., 2020). A BF greater than 1 indicates that the data provide more evidence in favor of the first hypothesis. In contrast, a BF less than 1 shows that the data provide more evidence in favor of the second hypothesis. A BF of 1 indicates that the data offer equal evidence for both hypotheses. For example, a BF_{01}= 4.5 tells that the observed data are 4.5 times more likely to occur under the null hypothesis.

To interpret BF, it is possible to use a scale such as Jeffreys' scale (Jeffreys, 1961), where a value greater than 1 to 3 is considered “anecdotal” evidence, between 3 and 10 is considered “moderate” evidence and greater than 10 is considered “strong” evidence.

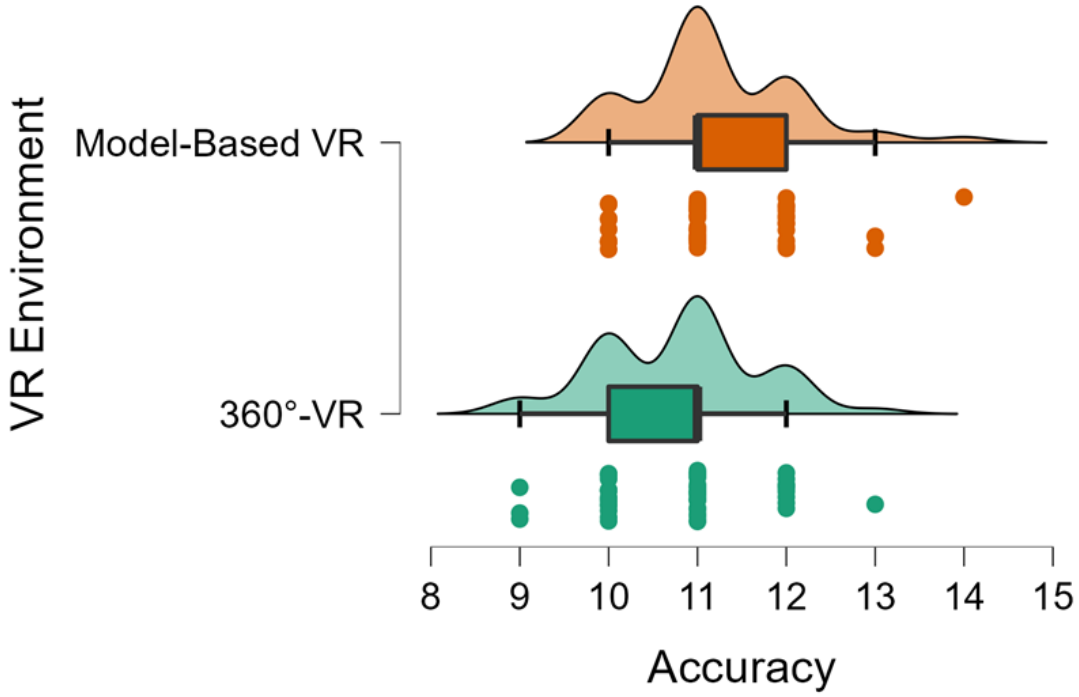
In this study, data pre-processing was performed using R, while the analyses were run using JASP 0.16.4.0 (JASP Team, 2022). The main measures which were contrasted and analyzed between the groups concerned were Task Performance (Accuracy and Time), questionnaire scores (ITQ, ITC-SOPI scales, SUS) and Correlations between them. Unless otherwise specified, each analysis was conducted using the following Bayesian framework provided by JASP software. For independent samples, test-related statistics, effect size δ was assigned to a Cauchy prior distribution with $r=1/\sqrt{2}$, while for correlations, the stretched beta prior width was set to 1. Additionally, for independent samples hypothesis testing, the null hypothesis postulated that there was no difference between groups (H_0: $\delta=0$). The alternative hypothesis H_1 assumed a difference but did not postulate the direction of the possible effect (H_1: $\delta\neq0$). For correlation, the hypothesis testing assessed whether the data were more likely to occur under H_0 (no linear association between the two variables) or under H_1 (there was an association between the two variables).

3.2.3 Results

3.2.3.1 Task Performance

Here the dependent variables of interest were Accuracy, considered as the number of detected targets, and the elapsed Time (in seconds) for task completion.

The descriptives and the corresponding raincloud plots showing the distribution of Accuracy performance for each experimental group are reported in the Figure 18.



Descriptives

	Group	N	Mean	SD	SE	Coefficient of variation	95% Credible Interval Lower	95% Credible Interval Upper
Accuracy	360°-VR	50	10.800	0.881	0.125	0.082	10.550	11.050
	Model-Based VR	49	11.204	0.866	0.124	0.077	10.955	11.453

Figure 18. Raincloud plots show the Accuracy scores distribution in the two groups: Model-Based VR and 360°-VR. The descriptives table is reported under the two plots. Images created with JASP.

Considering that the correct number of targets was 11, participants showed a high mean level of accuracy in both conditions (i.e., ceiling effect). In particular, the 360°-VR group reported a mean accuracy of 10.8 (SD = 0.88), while the Model-Based VR group showed a mean accuracy of 11.2 (SD=0.86). To determine if this difference was supported by evidence, non-parametric test statistics were run, namely the Bayesian Mann-Whitney U Test (Yuan and Johnson, 2008). The Figure 19 shows that there is only anecdotal evidence supporting H_1. In particular, BF_{10}=1.35 suggest that data appear to be only 1.35 times more likely under H_1 than under H_0. Thus, it is impossible to draw conclusions supporting an existing difference in accuracy performance between groups.

Target Selection Task Accuracy

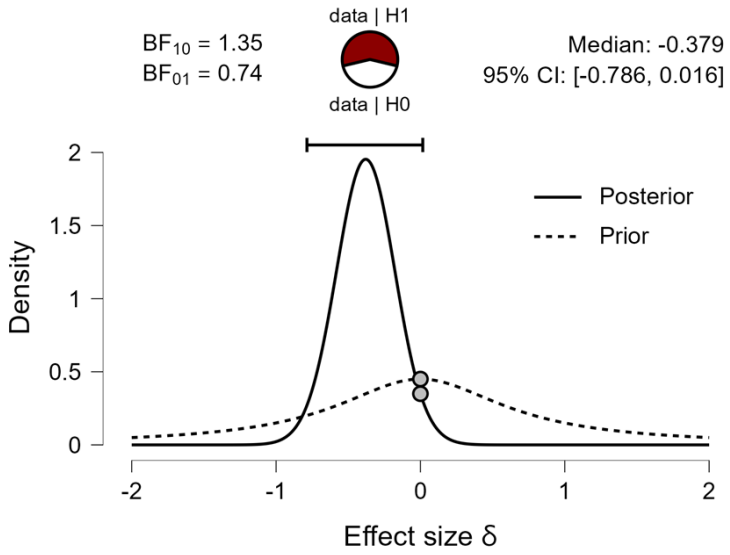
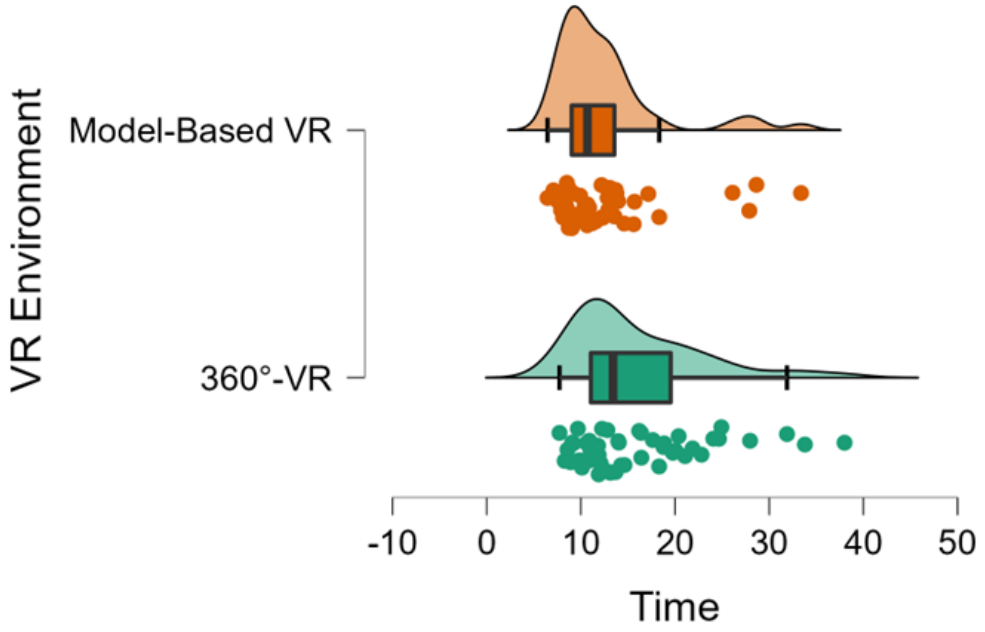


Figure 19. Prior and Posterior plot reports the Bayesian Mann-Whitney U Test results for the parameter δ (Accuracy, Target Selection Task). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The median and 95% central credible interval of the posterior distribution are displayed in the top right corner. Image created with JASP.

Regarding Time performance, it was log-transformed due to issues with non-normality. The descriptives and the corresponding raincloud plots showing the distribution of the Time each experimental group is reported in the Figure 20.



Descriptives

	Group	N	Mean	SD	SE	Coefficient of variation	95% Credible Interval Lower	95% Credible Interval Upper
Time	360°-VR	50	15.891	6.920	0.979	0.435	13.924	17.857
	Model-Based VR	49	12.470	5.709	0.816	0.458	10.830	14.110
LogTime	360°-VR	50	2.686	0.391	0.055	0.145	2.575	2.797
	Model-Based VR	49	2.449	0.366	0.052	0.149	2.343	2.554

Figure 20. A raincloud plot showing the distribution of Time scores in the two groups: Model-Based VR and 360°-VR. The descriptives table is reported under the plot and also shows the descriptives data concerning LogTime. Images created with JASP.

The mean elapsed seconds for task execution was 12.47s (SD=5.7) for the 360°-VR group and 15.89s (SD=6.92) for the Model-Based VR group.

A Bayesian Independent-Samples T-Test (Rouder et al., 2009) was run on the LogTime transformation data to determine if this difference is statistically meaningful.

Results support H_1: in particular, the resulting BF_{10}=16.313 suggests that data appear to be 16.313 times more likely under H_1 than under H_0. The Figure 21 shows that for LogTime, the median of the resulting posterior distribution for δ equals 0.587 with a central 95% credible interval for δ that ranges from 0.191 to 0.990.

Target Selection Task

LogTime

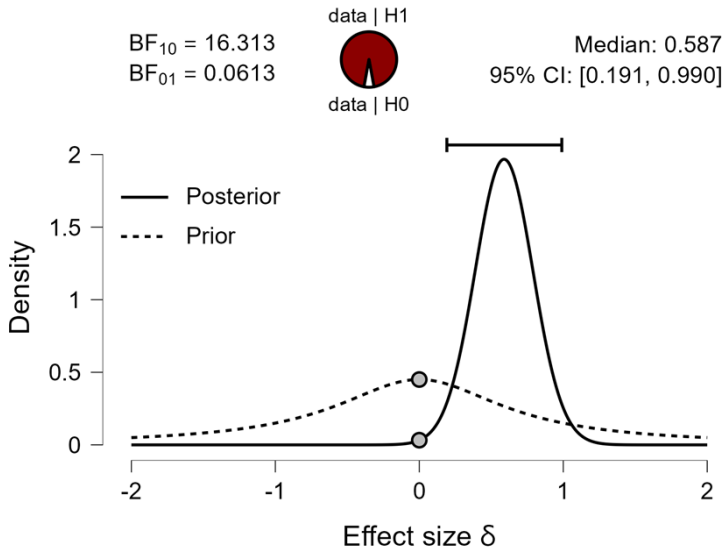


Figure 21. Prior and Posterior plot reports the Bayesian Independent Samples T-Test results for the parameter δ (LogTime, Target Selection Task). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The median and 95% central credible interval of the posterior distribution are displayed in the top right corner. Image created with JASP.

3.2.3.2 User-Experience

Questionnaires' scores which were included in the analysis, involved: ITQ total score, all four ITC-SOPI dimensions and SUS total score.

The descriptives concerning each of these questionnaire's scores are presented in Table 2.

Descriptives

	Group	N	Mean	SD	SE	Coefficient of variation	95% Credible Interval Lower	95% Credible Interval Upper
ITQ	360°-VR	50	4.242	0.780	0.110	0.184	4.020	4.464
	Model-Based VR	49	4.325	0.670	0.096	0.155	4.133	4.518
ITC_ES	360°-VR	50	3.816	0.401	0.057	0.105	3.702	3.930
	Model-Based VR	49	3.924	0.427	0.061	0.109	3.801	4.046
ITC_NE	360°-VR	50	1.907	0.708	0.100	0.371	1.705	2.108
	Model-Based VR	49	1.871	0.909	0.130	0.486	1.610	2.132
ITC_SP	360°-VR	50	3.158	0.500	0.071	0.158	3.016	3.300
	Model-Based VR	49	3.207	0.508	0.073	0.158	3.061	3.353
ITC_EV	360°-VR	50	3.820	0.576	0.081	0.151	3.656	3.984
	Model-Based VR	49	3.576	0.748	0.107	0.209	3.361	3.790
SUS_T	360°-VR	50	72.700	10.532	1.489	0.145	69.707	75.693
	Model-Based VR	49	76.684	16.484	2.355	0.215	71.949	81.418

Table 2. Table showing the descriptives data for each analysed questionnaire (Target Selection Task). Abbreviations. ITQ: Immersive Tendencies Questionnaire; ITC_ES: ITC-SOPI Engagement scale; ITC_NE: ITC-SOPI Negative Effects scale; ITC_SP: ITC-SOPI Spatial Presence scale; ITC_EV: ITC-SOPI Ecological Validity scale; SUS_T: System Usability Scale Total Score. Table created with JASP.

Results concerning Bayesian Mann-Whitney U test statistics are reported in Table 3.

Starting with the ITQ, results supported the absence of differences in self-report measures of immersive tendencies between groups. In particular, the observed data are 4.291 times more likely under H_0 (BF_{01}=4.291) (Fig. 22). Similar evidence supports the absence of differences between groups for all the scales of the ITC-SOPI, with evidence concerning the Ecological Validity scale (ITC_EV) being more ambiguous, with a BF_{01}= 1.246 (Fig. 23). The same uncertainty in supporting either H_0 or H_1 was also detected for SUS_T, with a BF_{01}=1.181 (Fig. 24).

Bayesian Mann-Whitney U Test ▼

	BF_{01}	W	Rhat
ITQ	4.291	1212.000	1.000
ITC_ES	1.728	945.000	1.000
ITC_NE	3.373	1363.500	1.000
ITC_SP	3.723	1111.000	1.000
ITC_EV	1.246	1470.500	1.000
SUS_T	1.181	951.500	1.000

Note. Result based on data augmentation algorithm with 5 chains of 10000 iterations.

Table 3. Table showing the Bayesian Mann-Whitney U test statistics and BFs resulting from the analysis of each questionnaire (Target Selection Task). Abbreviations. ITQ: Immersive Tendencies Questionnaire; ITC_ES: ITC-SOPI Engagement scale; ITC_NE: ITC-SOPI Negative Effects scale; ITC_SP: ITC-SOPI Spatial Presence scale; ITC_EV: ITC-SOPI Ecological Validity scale; SUS_T: System Usability Scale Total Score. Table created JASP.

Target Selection Task

ITQ

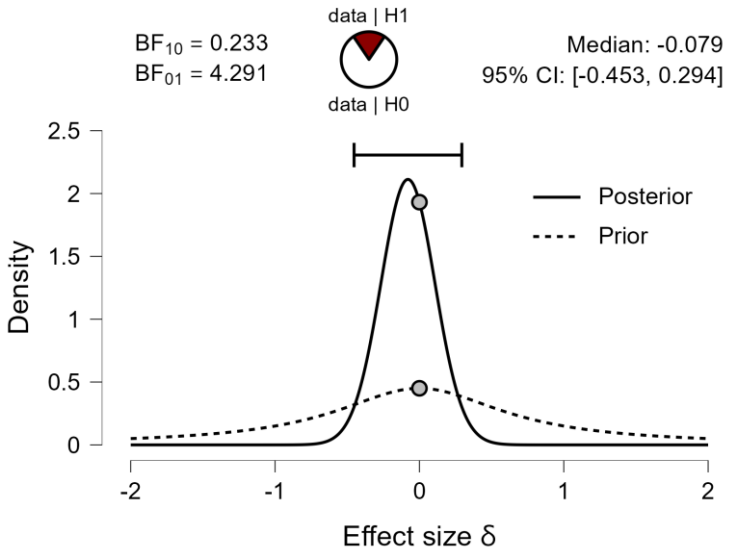


Figure 22. Prior and Posterior plot reports the Bayesian Mann-Whitney U Test results for the parameter δ (ITQ, Target Selection Task). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The median and 95% central credible interval of the posterior distribution are displayed in the top right corner. Abbreviations- ITQ: Immersive Tendencies Questionnaire. Image created with JASP.

Target Selection Task

ITC-EV

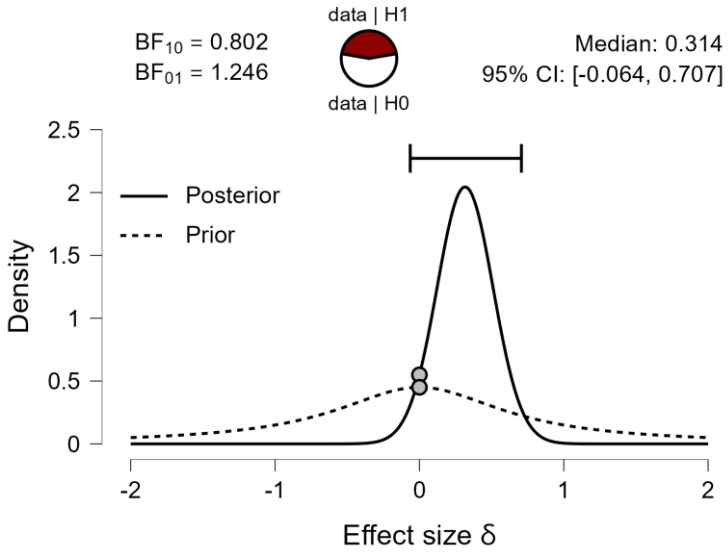


Figure 23. Prior and Posterior plot reports the Bayesian Mann-Whitney U Test results for the parameter δ (ITC-EV, Target Selection Task). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The median and 95% central credible interval of the posterior distribution are displayed in the top right corner. Abbreviations- ITC-EV: ITC-SOPI Ecological Validity scale. Image created with JASP.

Target Selection Task

SUS_T

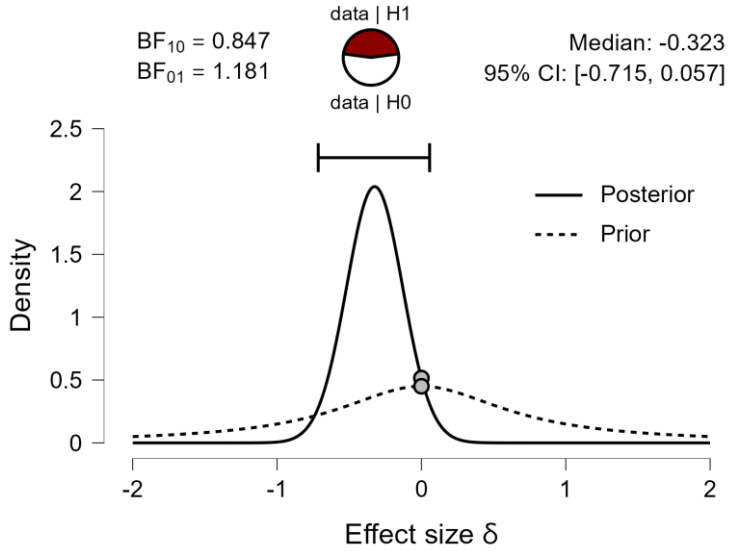


Figure 24. Prior and Posterior plot reports the Bayesian Mann-Whitney U Test results for the parameter δ (SUS_T, Target Selection Task). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The median and 95% central credible interval of the posterior distribution are displayed in the top right corner. Abbreviations- SUS_T: System Usability Score Total. Image created with JASP.

3.2.3.3 Correlations

In the last analysis, task performance measures (Accuracy, Time) were correlated with questionnaire scores (ITQ, ITC-SOPI scales, SUS). The results are reported in Table 4.

Here, the JASP output did not highlight any significant correlation. This is because the software flags correlations as significant only if the associated BF is at least greater than 10 (Goss-Sampson, 2020), which, considering Jeffreys' scale, defines the amount of evidence as "strong".

However, when considering Jeffreys' scale only, the Table shows that there are multiple pairs for which H_1 is possibly supported: that is, the negative correlation (r=-0.254) between LogTime and the Spatial Presence Scale of the ITC-SOPI (ITC_SP). In particular, the BF_{10}=2.973 suggests that

observed data are 2.973 more likely under H_1 (Fig. 26). The scatterplot reporting the correlation is presented in Figure 27.

Bayesian Pearson Correlations ▼

Variable		Accuracy	LogTime
1. Accuracy	Pearson's r	—	
	BF_{10}	—	
2. LogTime	Pearson's r	0.067	—
	BF_{10}	0.156	—
3. ITC_SP	Pearson's r	−0.067	−0.254
	BF_{10}	0.156	2.973
4. ITC_ES	Pearson's r	−0.135	−0.109
	BF_{10}	0.302	0.221
5. ITC_NE	Pearson's r	0.151	0.084
	BF_{10}	0.374	0.177
6. ITC_EV	Pearson's r	−0.139	−0.097
	BF_{10}	0.318	0.197
7. ITQ	Pearson's r	−0.009	−0.034
	BF_{10}	0.126	0.133
8. SUS_T	Pearson's r	−0.117	−0.184
	BF_{10}	0.243	0.649

* $BF_{10} > 10$, ** $BF_{10} > 30$, *** $BF_{10} > 100$

Table 4. Table showing the Bayesian Pearson Correlations between Performances and Questionnaires measures (Target Selection Task). Abbreviations. ITQ: Immersive Tendencies Questionnaire; ITC_ES: ITC-SOPI Engagement scale; ITC_NE: ITC-SOPI Negative Effects scale; ITC_SP: ITC-SOPI Spatial Presence scale; ITC_EV: ITC-SOPI Ecological Validity scale; SUS_T: System Usability Scale Total Score. Table created with JASP.

ITC_SP - LogTime

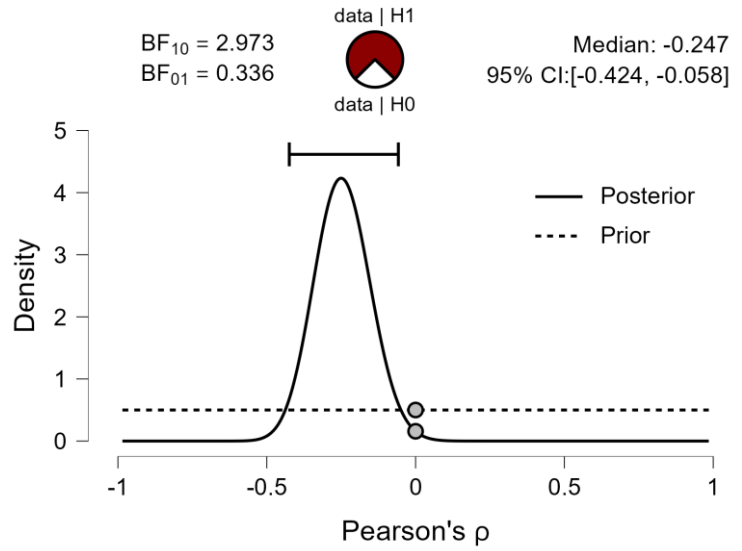


Figure 26. Prior and Posterior plot report the correlation test results for the parameter ρ (ITC_SP – LogTime, Target Selection Task). The reported analysis concerned the correlation between LogTime and ITC_SP. The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The median and 95% central credible interval of the posterior distribution are displayed in the top right corner. Abbreviations- ITC-SP: Spatial Presence scale of the ITC-SOPI. Image created with JASP.

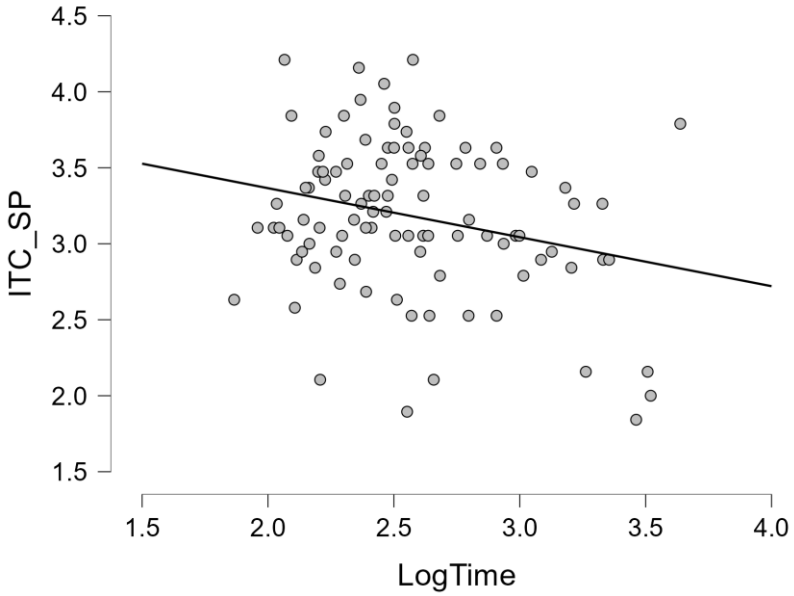


Figure 27. Scatterplot showing the correlation between the Spatial Presence scale of the ITC-SOPI (ITC_SP) and the LogTime measure. It is a negative correlation (r=-0.254): as the perceived sense of spatial presence augmented, the time spent to complete the task diminished.

3.2.4 Discussion

The current experiment examined how users' cognitive abilities and experience were affected by completing the same attentional task in two virtual variants - one based on 360°-VR and the other based on Model-Based VR. The study aimed to determine if the graphical elements of the two environments had a different impact on users' performance and sense of presence. Additionally, the research paid particular attention to how "natural" users perceived the environments, as measured by the Ecological Validity scale of the ITC-SOPI. This information is crucial in understanding which VE is best suited for developing specific neuropsychological tests in VR.

For what concerned performance, Accuracy and Time measures were collected and confronted between groups. Results showed no significant evidence for Accuracy in the two conditions. In particular, the measure did not distribute, showing a ceiling effect. This can be related to the fact that the TS task needed to be more challenging for the users. It is not a surprising result, as the participant of both groups were young and without cognitive issues. More interesting is the result related to time effects. On average, results showed that users exposed to the 360°-VR variant of the task were about 3.5 seconds slower when compared to the Model-Based VR group. This could be explained by the visual complexity which differed between the two environments. Even considering the best possible workflow, Model-Based VR is not yet ready to reach the photorealism provided by 360°-VR (Chalmers and Ferko al., 2010). Recreating a very detailed environment using Model-Based VR remains a challenge which results in very time consuming and requires a lot of human resources (e.g., a team of skilled programmers and 3D artists) (Nason et al., 2020). It is then possible that users were facilitated during Model-Based task completion because the stimuli were delivered in a less visually impacting environment, which was also characterized by lower distractive power.

Passing to user-experiences results, no differences were found between groups concerning individual immersive tendencies. This is a prerequisite for controlling the possible confounding effect of individual predisposition to immersion. Surprisingly, no evidence supported existing differences in

the dimensions investigated by the ITC-SOPI. However, the Ecological Validity scale results still need to be clarified, with the amount of evidence only slightly propending for H_0. Finally, there were no differences in the perceived usability between groups: this controls the potential confounding effect, which could have derived from an application being considered more usable than the other. This result is consistent with the fact that, despite different VR environments, the two applications shared a common setup and interaction paradigm for the administration.

Moreover, no significant correlations were found between performance and user-related experience between groups. The only small tendency which was detected concerned a possible negative correlation between LogTime and ITC_SP (r=-0.254): if confirmed, this effect indicates that participants were more rapid in concluding the task if they felt more spatially present in the VE. This may suggest that a stronger perception of presence within the VE could have improved cognitive performance.

Considered globally, these results showed preliminary evidence suggesting that the kind of VE used for implementing a VR cognitive task may play a role and thus need to be carefully considered. However, there is insufficient evidence supporting the use of 360°-VR over Model-Based VR or vice versa. A possible source of confusion in the present study could be found in the intrinsic nature of the task used for studying possible differences. In fact, the target selection required the users to focus their attention on searching for the stimuli without paying too much attention to the environmental characteristics. Moreover, on average, the low amount of time spent completing the task resulted in reduced exposure to the VEs, limiting presence-related feelings.

To disambiguate these results, an additional experiment was run. The research design was the same, which characterized Experiment One (between-subject, 2 conditions: 360°-VR vs Model-Based VR), but the task used for gathering additional evidence was different. Experiment Two included an immersive memory task, namely a FVE task, immediately followed by a delayed recall of the named

objects. Given the considerations from analysing the results obtained in the first experiment, the task involved a more extended exposure time to both VEs.

3.3 Experiment Two: Memory Task

3.3.1 Methods

3.3.1.1 Participants

In this additional experiment, 99 participants were recruited through advertisements within the University of Milano-Bicocca and were randomly assigned to two groups, one for the 360°-VR condition and one for the Model-Based VR condition. The study was conducted in accordance with the Helsinki Declaration of 1975, as revised in 2013 and approved by the Ethics Committee of the Department of Psychology of Milano-Bicocca. The exclusion criteria for this study included impairments that could have made it unsafe or uncomfortable for participants to be exposed to VR, such as a severe internist, psychiatric, neurological impairments or a previous history of motion sickness symptoms. After the experiment rollout, the samples included in the statistical analysis consisted of a group of 50 participants (29 females, 21 males) for the 360°-VR condition and a group of 49 participants (35 females, 14 males) for the model-based condition. The mean age of participants within the 360°-VR group was 24.26 years (SD = 5.558, range = 19-48), and the mean years of education were 14.5 years (SD = 1.5, range = 13–16). The mean age of participants within the Model-based VR group was 24.1 years (SD = 4.56, range = 19-47), and the mean years of education were 14.5 years (SD = 1.49, range = 13–16). All participants were provided with detailed information about the procedures, potential risks, and data usage, both verbally and in written form. After being fully informed, participants were asked to provide written consent to participate in the study.

3.3.1.2 Application Design and Task Administration

The memory task was implemented in two versions -360°VR and Model-Based - with the same modalities and tools used for the previous attentional task (see Paragraph 3.2.2.2). The familiarization

phase was also identical. In the test phase, participants found themselves in the center of the same living room. They needed to explore the virtual scenario and denominate all the objects they could recognize in a timeframe of three minutes. After the exposition, they were asked to recall all the named objects immediately. The experimenter manually recorded all the users' responses, namely the number of reported things and the time spent performing the recall (in seconds). Finally, they were asked to fill out the user-experience questionnaires. A comparison showing the graphical differences between the two variants of the task is presented in Figure 28.

3.3.1.3 Administered Questionnaires

The questionnaires used were the same utilized for the previous attentional task (see Paragraph 3.2.2.3).



Figure 28. A comparison of the Free Visual Exploration task between the two modalities. The left image shows the 360°-VR variant of the task, while the right image shows the Model-Based variant of the task.

3.3.1.4 Experimental Procedure

The experimental procedure used was the same as the previous attentional task (see Paragraph 3.2.2.4).

3.3.1.5 Statistical Analysis

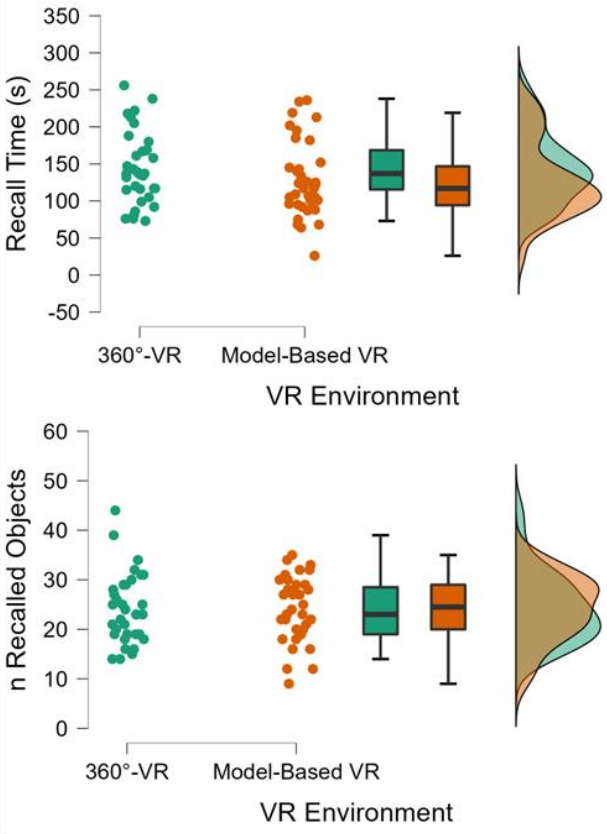
The statistical analyses were the same utilized for the previous attentional task and shared the same assumptions (see Paragraph 3.2.2.5).

3.3.2 Results

3.3.2.1 Task Performance

Here the dependent variables of interest were the number of Recalled Object and the Recall Time (in seconds).

The descriptives and the corresponding raincloud plots showing the distribution of both Recall-related performances for each experimental group are reported in the Figure 29.



Descriptives

	Group	N	Mean	SD	SE	Coefficient of variation	95% Credible Interval	
							Lower	Upper
Recall Time (s)	360°-VR	35	144.600	48.158	8.140	0.333	128.057	161.143
	Model-Based VR	36	127.444	51.287	8.548	0.402	110.091	144.798
n Recalled Objects	360°-VR	50	23.740	6.282	0.888	0.265	21.955	25.525
	Model-Based VR	49	24.490	6.285	0.898	0.257	22.684	26.295

Figure 29. Raincloud plots show the Recall Times distribution and the number of Recalled Object in the two groups: Model-Based VR and 360°-VR. The descriptives table is reported under the two plots. Images created with JASP.

On average, participants exposed to the 360°-VR condition needed 144.6 seconds (SD=48.15) to recall all the named objects. On the other hand, the participants exposed to the Model-Based VR variant of the task required a mean of 127.44 seconds (SD=51.28) for recalling the named objects.

A Bayesian Independent-Samples T-Test was run to test the evidence supporting H_0 and H_1. The amount of collected evidence does not clearly support one hypothesis over the other, with a BF_{01}=1.664 suggesting that the observed data are only 1.644 times more likely under H_0 (Fig. 30).

Free Visual Exploration Task

Recall Time

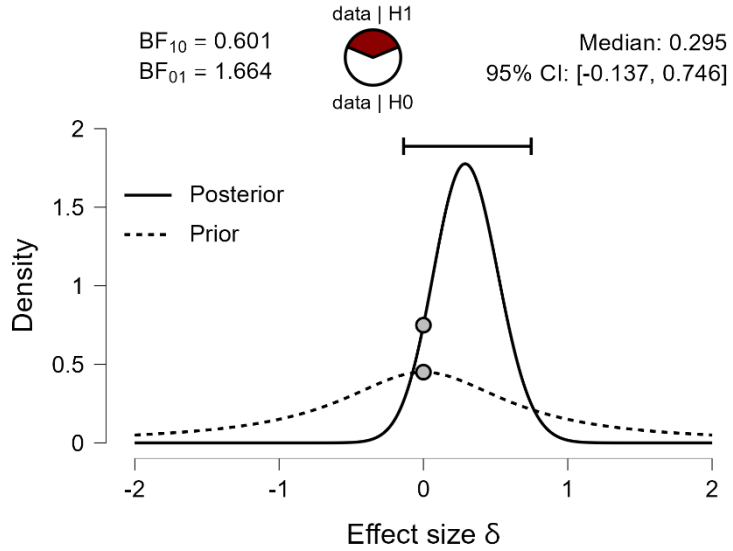


Figure 30. Prior and Posterior plot reports the Independent Samples T-Test results for the parameter δ (Recall Time, Free Visual Exploration Task). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The median and 95% central credible interval of the posterior distribution are displayed in the top right corner. Total. Image created with JASP.

For the number of Recalled Objects, a Bayesian Independent Samples T-Test showed moderate evidence supporting H_0. In particular, a BF_{01}=4.037 supports H_0 as 4 times more likely after observing the data (Fig. 31).

Free Visual Exploration Task

n Recalled Objects

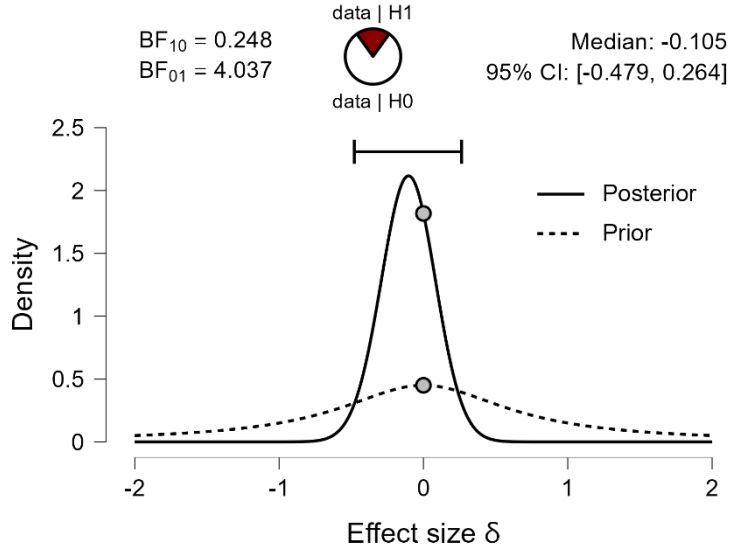


Figure 31. Prior and Posterior plot reporting the results for the Independent Samples T-Test for the parameter δ (n Recalled Objects, Free Visual Exploration Task). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The median and 95% central credible interval of the posterior distribution are displayed in the top right corner. Total. Image created with JASP.

3.3.2.2 User-Experience

The descriptive data concerning the self-report measures obtained through questionnaires administration are presented in Table 5.

Descriptives

	Group	N	Mean	SD	SE	Coefficient of variation	95% Credible Interval Lower	95% Credible Interval Upper
ITQ	360°-VR	50	4.292	0.771	0.109	0.180	4.073	4.511
	Model-Based VR	49	4.228	0.706	0.101	0.167	4.026	4.431
ITC_ES	360°-VR	50	3.627	0.385	0.054	0.106	3.517	3.736
	Model-Based VR	49	3.599	0.374	0.053	0.104	3.491	3.706
ITC_NE	360°-VR	50	1.917	0.657	0.093	0.343	1.730	2.104
	Model-Based VR	49	1.969	0.718	0.103	0.365	1.763	2.176
ITC_SP	360°-VR	50	3.009	0.490	0.069	0.163	2.870	3.149
	Model-Based VR	49	3.105	0.511	0.073	0.164	2.959	3.252
ITC_EV	360°-VR	50	3.912	0.502	0.071	0.128	3.769	4.055
	Model-Based VR	49	3.571	0.579	0.083	0.162	3.405	3.738
SUS_T	360°-VR	50	75.750	9.109	1.288	0.120	73.161	78.339
	Model-Based VR	49	75.051	12.038	1.720	0.160	71.593	78.509

Table 5. Table showing the descriptives data for each analysed questionnaire (Free Visual Exploration task). Abbreviations. ITQ: Immersive Tendencies Questionnaire; ITC_ES: ITC-SOPI Engagement scale; ITC_NE: ITC-SOPI Negative Effects scale; ITC_SP: ITC-SOPI Spatial Presence scale; ITC_EV: ITC-SOPI Ecological Validity scale; SUS_T: System Usability Scale Total Score. Table created with JASP.

The results of Bayesian Mann-Whitney U test statistics are summarized in Table 6.

Bayesian Mann-Whitney U Test ▼

	BF_{10}	W	Rhat
ITQ	0.231	1266.000	1.000
ITC_ES	0.240	1345.000	1.000
ITC_NE	0.226	1208.500	1.000
ITC_SP	0.321	1055.000	1.000
ITC_EV	8.015	1637.500	1.000
SUS_T	0.221	1231.500	1.000

Note. Result based on data augmentation algorithm with 5 chains of 10000 iterations.

Table 6. Table showing the Bayesian Mann-Whitney U test statistics and BFs resulting from the analysis of each questionnaire (Free Visual Exploration Task). Abbreviations. ITQ: Immersive Tendencies Questionnaire; ITC_ES: ITC-SOPI Engagement scale; ITC_NE: ITC-SOPI Negative Effects scale; ITC_SP: ITC-SOPI Spatial Presence scale; ITC_EV: ITC-SOPI Ecological Validity scale; SUS_T: System Usability Scale Total Score. Table created JASP.

Starting with the ITQ, there is moderate evidence supporting H_0: the observed data appeared 4.336 times more likely under H_0 (Fig. 32).

Free Visual Exploration Task

ITQ

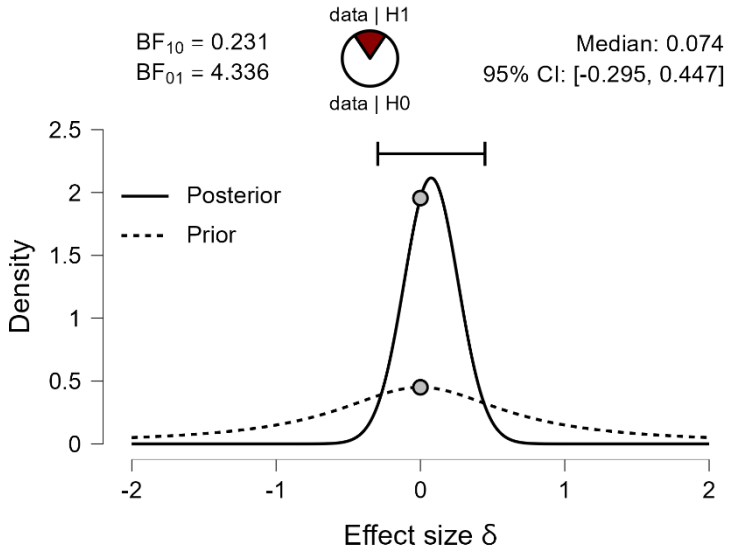


Figure 32. Prior and Posterior plot reports the Bayesian Mann-Whitney U Test results for the parameter δ (ITQ, Free Visual Exploration Task). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The median and 95% central credible interval of the posterior distribution are displayed in the top right corner. Abbreviations- ITQ: Immersive Tendencies Questionnaire. Image created with JASP.

Concerning the scales ITC-SOPI, there is moderate evidence supporting H_1 for the Ecological Validity scale. Namely. H_1 appeared eight times more likely than H_0 after observing the data (Fig. 33).

For parameter estimation, δ was assigned a Cauchy prior distribution with $r=1/\sqrt{2}$. The figure shows that for ITC_EV, the median of the resulting posterior distribution for δ equals 0.527 with a central 95% credible interval for δ that ranges from 0.139 to 0.929.

No differences were detected for the other subscales.

Free Visual Exploration Task

ITC-EV

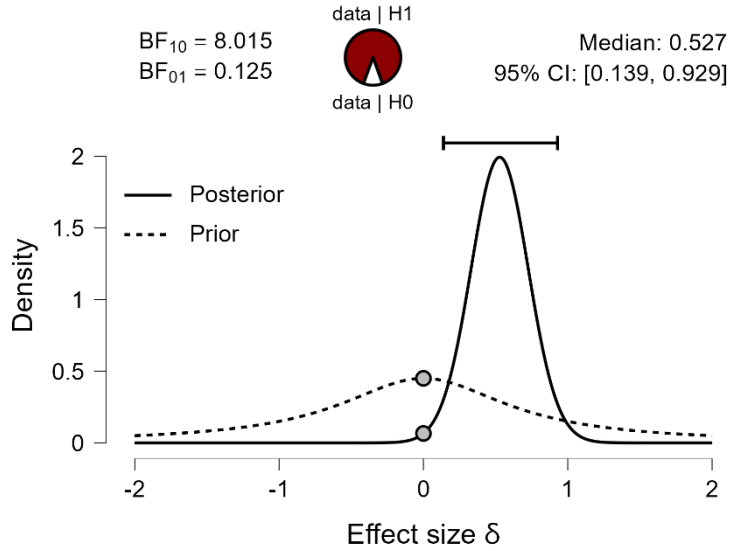


Figure 33. Prior and Posterior plot reports the Bayesian Mann-Whitney U Test results for the parameter δ (ITC_EV, Free Visual Exploration Task). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The median and 95% central credible interval of the posterior distribution are displayed in the top right corner. Abbreviations- ITC-EV: ITC-SOPI Ecological Validity scale. Image created with JASP.

Lastly, no differences were found between groups in the perception of usability. Indeed, SUS_T has a moderate amount of evidence supporting H_0, which appears 4.533 more likely than H_1 (Fig. 34).

Free Visual Exploration Task

SUS_T

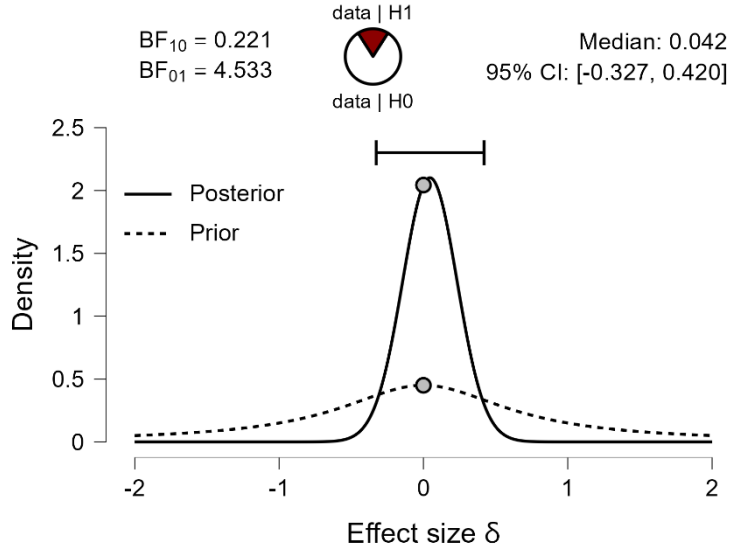


Figure 34. Prior and Posterior plot reports the Bayesian Mann-Whitney U Test results for the parameter δ (SUS_T, Free Visual Exploration Task). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The median and 95% central credible interval of the posterior distribution are displayed in the top right corner. Abbreviations- SUS_T: System Usability Score Total. Image created with JASP.

3.3.2.3 Correlations

In the last analysis, task performance measures (Recall Time and n Recalled Objects) were correlated with questionnaires scores (ITQ, ITC-SOPI scales, SUS). The results are reported in Table 7.

Bayesian Pearson Correlations ▼

Variable		Recall Time (s)	n Recalled Objects
1. Recall Time (s)	Pearson's r	—	
	BF_{10}	—	
2. n Recalled Objects	Pearson's r	0.373*	—
	BF_{10}	22.551	—
3. ITC_ES	Pearson's r	−0.059	−0.041
	BF_{10}	0.167	0.136
4. ITC_NE	Pearson's r	−0.061	0.014
	BF_{10}	0.168	0.127
5. ITC_SP	Pearson's r	−0.103	−0.034
	BF_{10}	0.211	0.133
6. ITC_EV	Pearson's r	−0.126	0.026
	BF_{10}	0.255	0.130

* $BF_{10} > 10$, ** $BF_{10} > 30$, *** $BF_{10} > 100$

Table 7. Table showing the Bayesian Pearson Correlations between Performances and Questionnaires measures (Free Visual Exploration Task). Abbreviations. ITQ: Immersive Tendencies Questionnaire; ITC_ES: ITC-SOPI Engagement scale; ITC_NE: ITC-SOPI Negative Effects scale; ITC_SP: ITC-SOPI Spatial Presence scale; ITC_EV: ITC-SOPI Ecological Validity scale; SUS_T: System Usability Scale Total Score. Table created with JASP.

A significant difference was found only in the correlation between the number of Recalled Objects and Recall Time (r=0.373). In particular, observed data appears 22.55 times more likely under H_1.

3.3.3 Discussion and Model Comparison

This experiment aimed to gather additional information and overcome the possible confounding effect related to the low time of exposition to VR, which characterized the previous TS Task. The shortness of the experience could have negatively affected presence-related feelings, including the perceived ecological validity in each environment. To disambiguate this possible confounder, a FVE task was implemented using 360°-VR and Model-Based VR. This task included a much longer exposition to the VR environments (i.e., 3 minutes), with performance measures defined by the number of recalled objects after the exposition and the associated time needed to complete the recall task. The administered questionnaires were the same for Experiment One: ITQ, ITC-SOPI and SUS.

Different from Experiment One, in Experiment Two, no differences were found between the two groups for what concerned performance, with only the results coming to the analysis of the Recall Time measure needing to be more precise, not propending for one hypothesis over another.

For what concerned questionnaires, no differences were found for the ITQ total score. Similarly to Experiment One, this result certificates that the two groups were comparable for individual predisposition to immersion, controlling this as a potential confounding effect which could have influenced the user's responses to the ITC-SOPI. The most interesting result came from this last questionnaire. The only significant difference was found in the Ecological Validity scale, with moderate evidence suggesting that, on average, participants considered more realistic the 360°-VR variant of the task. Additionally, considering SUS_T results, participants did not perceive one task variant as more usable. Lastly, no correlations were detected between the performance measures and the questionnaire scores. This indicates that the 360°-VR variant of the task was perceived as more ecological, but this does not mean that there was an influence on performance.

The results from both experiments suggest that the kind of VR used for implementing a specific cognitive task has a specific impact on the perceived ecological validity of the experience, with the type of task being more in the background. However, this hypothesis needed to be tested, as the results coming from Experiment One did not clearly show a difference in the perceived ecological validity between the two groups. To do that, data coming from both experiments were aggregated, and a Bayesian ANOVA was performed to contrast the predictive performance of competing models (Rouder et al., 2016). The predictors included in each model were the type of VR Environment (360°-VR and Model-Based VR) and the type of Task (Target Selection and Free Visual Search). JASP priors (r scale prior width = 0.5 for fixed effects, r scale prior width = 0.354 for covariates) were used.

The results are reported in Table 8. For the Ecological Validity scale, results showed that the observed data appeared more likely under the model which included only VR Environment as a predictor.

Ecological Validity ▾

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
VR Environment	0.200	0.796	15.615	1.000	
Task + VR Environment	0.200	0.142	0.660	0.178	0.940
Task + VR Environment + Task ✻ VR Environment	0.200	0.032	0.134	0.041	2.389
Null model	0.200	0.025	0.105	0.032	9.942e−4
Task	0.200	0.004	0.018	0.006	0.063

Tablet 8. Comparison of models based on the Ecological Validity scale in the two experiments. The table includes columns for the predictors used in each model (Models), the prior P(M) and posterior P(M|data) probabilities of the model (BF_M), the posterior odds of the model, and the Bayes factors for each model compared to the best model (BF_{10}). The final column estimates the numerical error in calculating the Bayes factor. The models are listed in order of lowest to highest Bayes factor in relation to the best model.

This indicates that the factor that influenced the perceived ecological validity within the experience relied on the type of VR used, not the kind of task implemented. Considered within the framework concerning the development of 3.0 neuropsychological tests, this result supports the use of 360°-VR as a tool for delivering highly realistic stimuli, which can enhance the ecological validity of the assessment procedure. Thus, when compared to Model-Based VR, 360°-VR presents a trade-off: it does not provide the same level of interaction, but, on the other hand, it can give a more realistic stimulation which can elicit cognitive processes in a way which is similar to real life. This allows the collection of more meaningful measures of cognitive functioning, which is fundamental to the diagnostic and rehabilitation process. The upcoming VR-based neuropsychological tests can consider the use of 360°-VR for implementing assessment tasks, trying to exploit the advantages coming from the higher ecological validity provided. At the same time, it is essential to find workarounds to surpass the limitations imposed by the lower interaction capabilities that characterize this kind of VR environment.

4. The Mental Status 360 ° (MS-360°) Developing a Virtual Reality Test for an Ecological Screening of Cognitive Functions

4.1 Theoretical Background

The empirical evidence from the studies presented in Chapter 3 shows that 360°-VR can provide a more ecological presentation of stimuli within VR-based neuropsychological tasks. This fundamental feature encourages the use of this kind of VR environment for developing novel tools, possibly counterbalancing its current limitations. In the field of clinical neuropsychology, the interaction limitations that characterize 360°-VR may indeed offer a safer and simpler setting for assessment. In fact, this technology provides only a fixed 3-DoF exploration (see Paragraph 1.2.2), allowing the users to join a still 360° view of the scenario without permitting physical movement. For example, these interaction limitations do not represent an issue in all those situations in which patients are affected by motor impairments. Then, if the task to implement is designed accordingly to the known characteristics of the target clinical population, these limitations can result in much less impactful on user experience. The challenge is to develop simulated tasks not influenced by these limitations but at the same time complex enough to elicit the cognitive function to test and exploit the ecological advantage provided by 360°-VR.

Until recently, most experimental and commercial VR-based neuropsychological tests were developed using Model-Based VR environments (See Chapter 2). Much fewer are the test which explored the use of immersive 360°-VR. Two studies investigated the use of this technology for the assessment of executive functions in two different clinical populations: Parkinson's disease (Serino et al., 2017) and multiple sclerosis (Realdon et al., 2019). In both studies, the 360°-VR tool successfully classified patients and controls, showing convergent validity with gold-standard neuropsychological tests. Additionally, user-experience self-reports showed how the assessment experience was considered engaging, improving the patient's compliance with the experience. The same preliminary results were found for what concerned memory assessment, with first Pieri et al.

(2021) and then Bruni et al., (2022) implementing a 360°-VR version of a Free Recall and Recognition task.

These studies showed that maybe 360-°VR reached a methodological maturity level within the neuropsychological assessment framework. Thus, there is enough evidence supporting this kind of VR for creating a tool allowing a more comprehensive assessment of cognitive functioning. As mentioned above, until now, 360°-VR were developed for assessing only single cognitive domains, namely memory and executive functions. The present study describes the development of a pilot 360°-VR neuropsychological test aimed at a first-level assessment of cognitive functioning. We named this test Mental Status 360° (MS-360°). This tool's main point of novelty is related to the number and diversity of tasks included within the simulation. First, a description of the application design and a brief presentation of each included task are provided. Then, the results from a between-group comparison of the scores obtained by participants in the MS-360° and a classical neuropsychological screening test are shown, followed by an analysis of user-experience.

4.2 Methods

4.2.1 Participants

In this between-subject study, two groups of participants were recruited: a group reporting Subjective Cognitive Impairment (SCI) (Reisberg et al., 2010; Stewart, 2012) and a group of healthy controls (HC). The firsts were recruited through voluntary access within the clinical activity of the CEMS Verona Memory Center ambulatory. At the same time, the seconds entered the experimentation thanks to contacts offered by the personnel of the clinical structure. The study was conducted in accordance with the Helsinki Declaration of 1975, as revised in 2013 and approved by the Ethics Committee of the Department of Psychology of Milano-Bicocca. The exclusion criteria for this study included an age < 60 years and impairments that could have made it unsafe or uncomfortable for participants to be exposed to VR, such as a severe internist, psychiatric, neurological impairments or having a previous history of motion sickness symptoms. The final samples included 20 participants

(9 females, 11 males) for the SCI group and 27 participants (14 females and 13 males) for the HC group. The first presented a mean age of 73.9 years (SD=6.66, range = 61-85) and a mean education of 11.2 years (SD=6.72, range = 3-24). The second presented a mean age of 65.7 years (SD=4.8, range = 60-76) and a mean education of 14.4 years (SD=2.46, range = 8-19). All participants were provided with detailed information about the procedures, potential risks, and data usage, both verbally and in written form. After being fully informed, participants were asked to provide written consent to participate in the study.

4.2.2 Application Design and Task Administration

The paradigm used for developing the MS-360° is similar to the one presented in Paragraph 3.2.2.2, with some differences related to the more complex nature of the included tasks and the target population of users. Similarly, the application consists of two different phases: a familiarization phase and a test phase. The first allows the participant to become familiar with the raycast interaction system and the 3-DoF exploration. In particular, the presentation of the starting screen is accompanied by audio instructions which tell the user how to point and select a button within the VE, labelled as “Try” (i.e., “Prova”) (Fig. 35a). To keep the instructions standardized, the audios were synthesized using Microsoft Speech API included in the Balabolka software, but with voice parameters modified (e.g., a slower pace) to be more understandable for older and possibly cognitively impaired users.

The familiarization task requires participants to find and select three different buttons showing the numbers “1”, “2”, and “3” (Fig. 35b). When selected, each of these buttons presents an audio instruction telling the users to find the following number. These icons are positioned on the right, behind, and left of the starting position. This task aims to introduce the users to understand how to explore the upcoming scenarios. Once the last button is selected, the starting scene of the test phase is loaded.



Figure 35. (a) A screenshot showing the starting screen of the MS-360°. The instructions (both audio and written) tell the user to point to the raycast and press the trigger button to select interactive buttons around the scene. (b) A screenshot showing how the buttons, including icon numbers, appeared to the user.

The test phase includes 14 scenarios, each presenting a specific cognitive task. Each scenario was developed, taking inspiration from classical neuropsychological tests but following a function-led approach for implementation (see Paragraph 2.2). The structure of the presentation is the same across all tasks. A first scenario starts reporting audio instructions for the upcoming task, with two buttons allowing the user to proceed with the test or to listen again to the audio instructions (Fig. 36a). A second scenario includes the actual task to perform, eventually followed by an audio prompt asking the users to tell their response, if needed, by task design. To guarantee a more comfortable experience when passing from one scenario to another, a loading screen with a fixed 5-second load time was implemented (Fig. 36b). The participants performed the passage from one scenario to the other using buttons placed in the scene.

The application was developed using the Unity game engine with the Oculus Integration SDK Version 29.0, and the HMD used for administration was the Meta Quest 2 (see Paragraph 3.2.2.2).

Each task included in the MS-360° is based on 360° spherical photos or videos captured using an Insta360 One X camera, all synchronized at a resolution of 5.7k (5760x2880 pixels) and post-processed using Insta360Studio. All videos run at 30 FPS: however, being all static videos, the chances of a possible occurrence of sickness symptoms are meagre. For each programmed task, these contents were imported in Unity and were used to fill the internal face of a sphere object using a custom shader C# script, positioning the camera origin at the center of this sphere. To provide a smooth delivery of the VR experience while keeping the setup portable, the MS-360° application was tethered to a high-performance MSI Stealth GS66 notebook with an Intel i7 CPU, NVIDIA RTX 2070 Max-Q GPU and 16 GB of RAM.

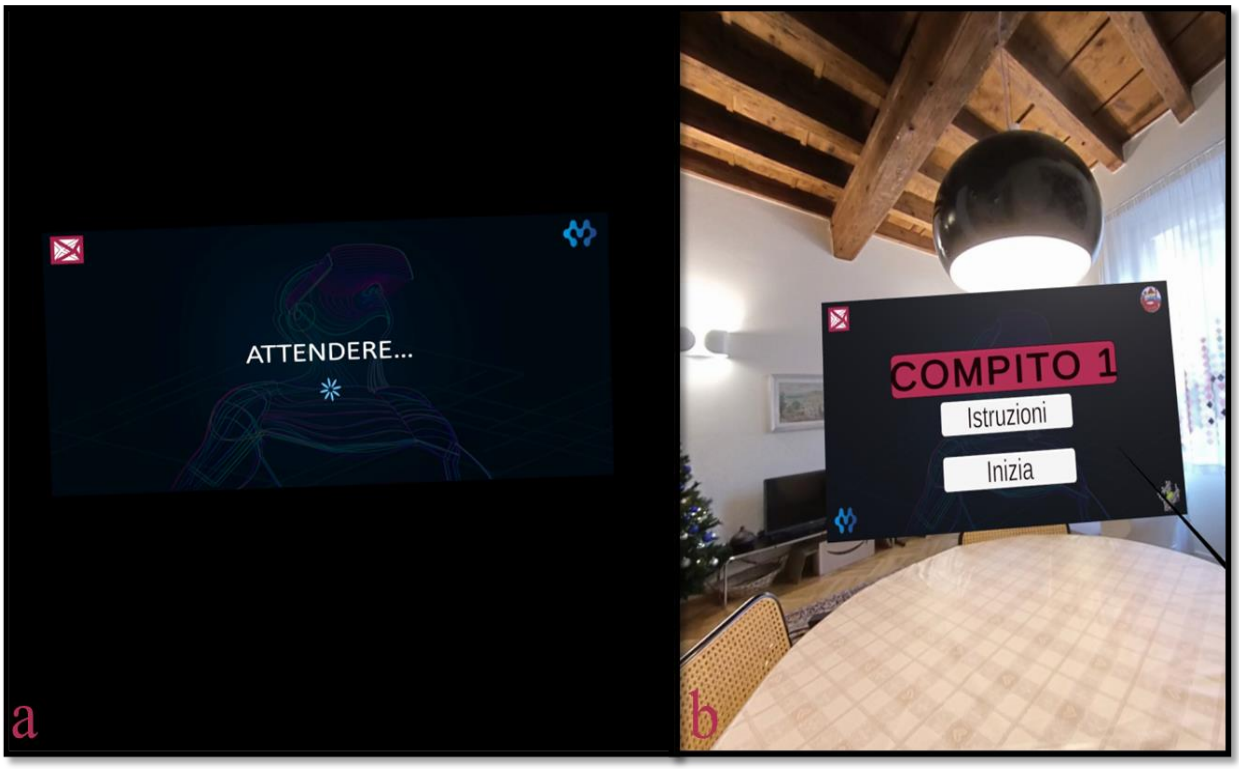


Figure 36. (a) A screenshot showing the 5-second loading screen of the MS-360°. (b) A screenshot showing how the pre-test phase was implemented for each task. The user can decide to proceed with the task execution or listen to the provided audio instructions.

The list of all 14 tasks and the related characteristics is reported in Table 9, while screenshots exemplifying each task are reported in the Appendix.

TASK NAME	DESCRIPTION	MAIN COGNITIVE FUNCTION
1. Free Visual Search	Search the living room and name all the recognized objects in 1 minute.	**-Visuospatial Functions**
2. Target Selection	Find and count 11 target plastic cups placed around the living room among distractors.	**-Attention**
3. Object Naming	Name 10 everyday objects placed over a table in a living room.	**-Language**
4. Object Recall (I)	Follow a TV transmission presenting a recipe and recall the given ingredients.	**-Memory**
5. Story Recall (I)	Follow a TV transmission presenting a news report and recall the story's details.	**-Memory**
6. Word Reading	Read a grocery list including 6 everyday objects.	**-Language**
7. Semantic Fluency	Tell names of fruits to add to the grocery list in one minute.	**-Language**
8. Cognitive Estimates	Estimate quantities concerning the objects on the grocery list.	**-Executive Functions**
9. Action Planning	Plan an action using the suitable object of the grocery list.	**-Executive Functions**
10. Written Comprehension	Read and interpret the title of a newspaper.	**-Language**
11. Sustained Attention	Follow a TV transmission presenting a lottery extraction and check if the extracted numbers are in the ticket.	**-Attention**
12. Object Recognition	Check a series of 10 boxes to find whether one or more objects named during Task 3 are present.	**-Memory**
13. Object Recall (D)	Recall the ingredients of the recipe presented during Task 4.	**-Memory**
14. Story Recall (D)	Recall the detail of the news report presented during Task 5.	**-Memory**

Table 9. The table reports the list of all the tasks included in the MS-360°. Notably, each of them is associated with the assessment of a cognitive function expected to be the most implied in the execution of that specific task. Abbreviations. (I): Immediate; (D): Delayed.

4.2.3 Administered Questionnaires

Questionnaires administered to participants aimed to collect data concerning dimensions related to perceived usability and cybersickness after MS-360° administration. For evaluating the first, the SUS (Brooke, 1986) was chosen (see Paragraph 3.2.2.3). The assessment of possible emerging symptoms related to motion sickness was performed using the Simulation Sickness Questionnaire (SSQ, Kennedy et al., 1993). This self-report questionnaire lists 16 sickness-related symptoms such as nausea, dizziness, headache, or eyestrain. After the exposition to simulation, participants need to rate how much they are affected by each listed symptom using a 4-point Likert scale ranging from "not at all" to "severe". The responses are then computed to obtain scores concerning the following sickness dimensions: Oculomotor, Disorientation, and Nausea. Moreover, a Total Severity score can be calculated.

4.2.4 Experimental Procedure

The experimental procedure included the Standard phase and the VR phase. The Montreal Cognitive Assessment (MOCA, Nasreddine et al., 2005) was administered during the standard phase. This gold-standard neuropsychological screening test measures various cognitive abilities like attention, memory, language, and visuospatial skills. It takes about 10-15 minutes to complete and includes tasks like immediate and delayed recall, attention and concentration, language, visual-spatial abilities, abstraction, and orientation. It is commonly used to detect early signs of cognitive decline and mild cognitive impairment in older adults. The VR phase included the administration of the MS-360°, followed by the compilation of SUS and SSQ questionnaires. The MS-360° administration was performed as follows and, on average, took about 20-25 minutes to complete. Participants were asked to sit comfortably on a swivel chair and wear the Meta Quest 2 headset. After that, the examiner launched the application from the connected notebook and followed the participants during the

execution of each task, seeing their POV through the notebook's screen (Fig. 37). All users' responses were recorded on a specifically designed paper protocol.

The order of the two experimental phases was counterbalanced within each group to avoid possible confounding effects.

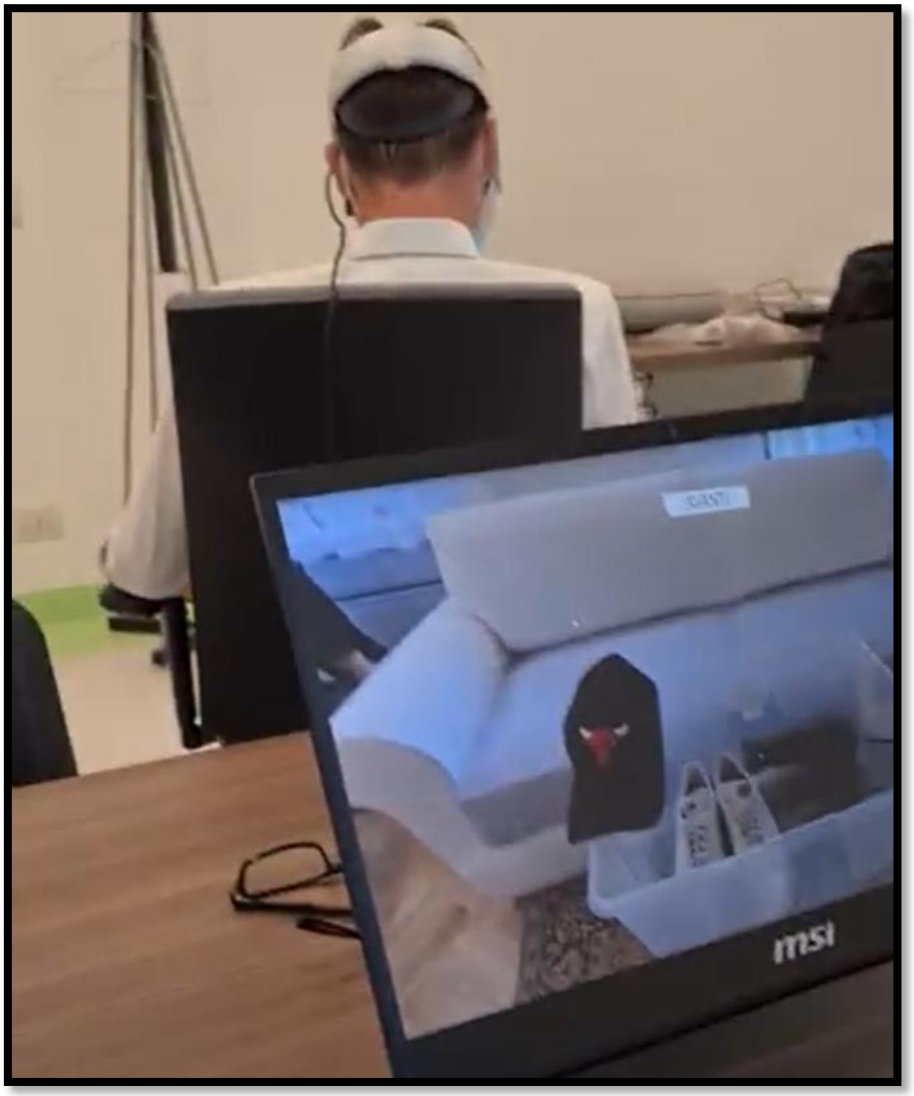


Figure 37. Example of the administration set in which the MS-360° was used. The experimenter followed the execution of the task during all the procedures through the screen of the notebook.

4.2.5 Statistical Analysis

The same Bayesian framework presented in Paragraph 3.2.2.5 was used for conducting the statistical analysis. Data pre-processing was performed using R, while the analyses were run using JASP 0.16.4.0 (JASP Team, 2022). The main measures which were analyzed between groups were Demographics (Age and Education), Test Scores (MOCA and MS-360° total scores), Questionnaire scores (SUS, SSQ) and Correlations between the different measures. The MS-360° Total Score is the sum of the accuracy scores derived from all 14 tasks. For each of them, a cut-off is specified to label the task as "not accurate" (0) or "accurate" (1). The value of each cut-off was initially set on a theoretical basis and similar tasks in the literature. This was then adjusted by checking the behaviour of the control group. To do so, the value of the 5^{th} percentile of the control group was studied, and this value was checked if it aligned with theoretical choices. If not, further theoretical considerations were taken to adjust the cut-off, which remains theoretically based in the end, although with some statistical considerations. It is worth noting that cut-offs of the subtests were only set to consider the latter as accurate or not, and did not contribute to the final global score. Thus, MS-360° Total Score could range from 0 to 14, while the MOCA's one from 0 to 30.

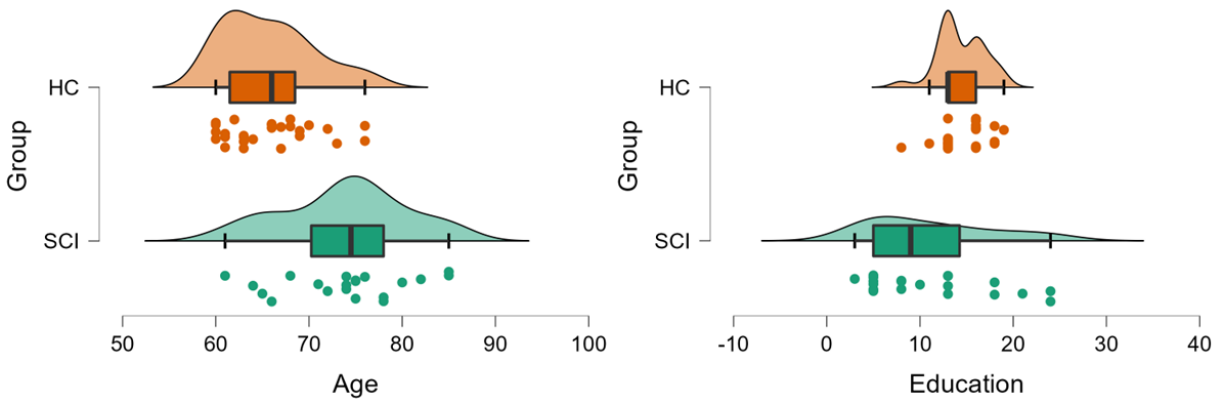
4.3 Results

4.3.1 Demographics

The dependent variables of interest were Age and Education. The descriptives data concerning these two variables are presented in Figure 38.

On average, participants in the SCI group presented a higher mean Age. This difference is supported by solid evidence, as the Bayesian Mann-Whitney U Test statistic reported a BF_{10}=23.8 (Fig. 39). The median of the resulting posterior distribution for δ equals 0.958 with a central 95% credible interval for δ that ranges from 0.328 to 1.619.

The same inference cannot be made for Education, as the resulting BF_{10}=1.14 does not provide enough evidence to support either H_0 or H_1 (Fig. 40).



Descriptive Statistics

	Age		Education	
	SCI	HC	SCI	HC
Valid	20	27	20	27
Missing	0	0	0	0
Mean	73.900	65.704	11.200	14.407
Std. Deviation	6.664	4.810	6.725	2.469
Minimum	61.000	60.000	3.000	8.000
Maximum	85.000	76.000	24.000	19.000

Figure 38. Raincloud plots show Age and Education data distribution in the two groups: Healthy Controls (HC) and Subjective Cognitive Impairment (SCI). The descriptives table is reported under the two plots. Images created with JASP.

Age

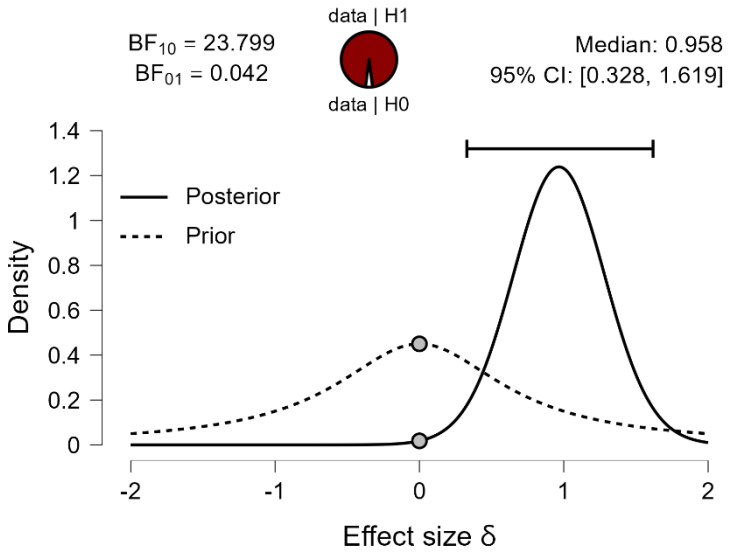


Figure 39. Prior and Posterior plot reports the Bayesian Mann-Whitney U Test results for the parameter δ (Age). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The posterior distribution's median and 95% central credible interval is displayed in the top right corner. Image created with JASP.

Education

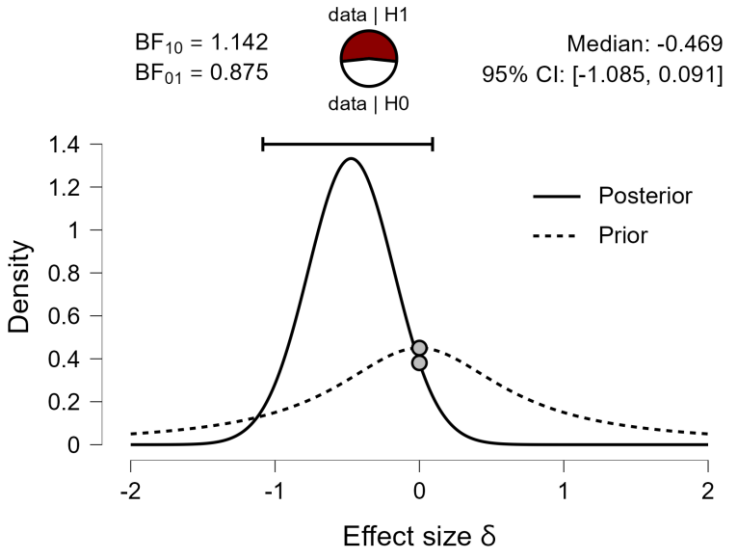


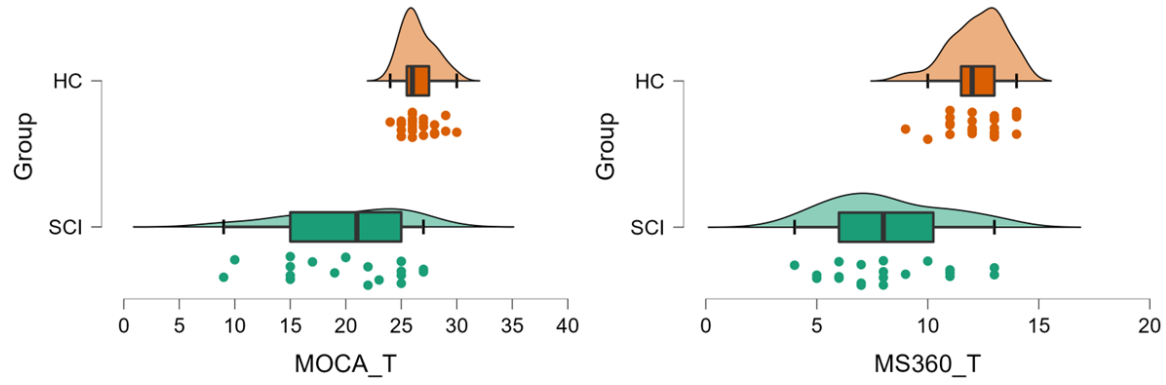
Figure 40. Prior and Posterior plot reports the Bayesian Mann-Whitney U Test results for the parameter δ (Education). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The posterior distribution's median and 95% central credible interval is displayed in the top right corner. Image created with JASP.

4.3.2 Test Performance

Here the dependent variables of interest were the MS-360° and MOCA total scores obtained by the two groups of participants. The descriptives data concerning these two variables are presented in Figure 41.

HC group reported a mean MOCA total score of 26.5 (SD=1.477, range= 24-30) and a mean MS-360° total score of 12.26 (SD=1.26, range= 9-14). Considering that the maximum reachable score for the first test was 30 and 14 for the second test, HC participants average presented high scores (i.e., ceiling effect). On the other hand, SCI participants reported a mean MOCA total score of 20 (SD=5.5, range= 9-27) and a mean MS-360 total score of 8.15 (SD=2.62, range= 4-13).

To test the statistical meaningfulness of these differences, multiple Bayesian Mann-Whitney U Tests were performed.



Descriptive Statistics ▼

	MOCA_T		MS360_T	
	SCI	HC	SCI	HC
Valid	20	27	20	27
Missing	0	0	0	0
Mean	20.050	26.519	8.150	12.259
Std. Deviation	5.482	1.477	2.621	1.259
Minimum	9.000	24.000	4.000	9.000
Maximum	27.000	30.000	13.000	14.000

Figure 41. Raincloud plots show MOCA (MOCA_T) and MS-360° (MS360_T) total scores distribution in the two groups: Healthy Controls (HC) and Subjective Cognitive Impairment (SCI). The descriptives table is reported under the two plots. Images created with JASP.

Starting with the MOCA results, the HC group showed a better performance when compared to the SCI group. This difference is supported by an extreme amount of evidence, as the reported BF_{10}=183 suggests that the observed data are 183 times more likely under H_1 than under H_0 (Fig. 42). The median of the resulting posterior distribution for δ equals -1.181 with a central 95% credible interval for δ that ranges from -1.836 to -0.532.

An even more substantial amount of evidence supports the statistical significance of the better performance shown by the HC group in the MS-360°. Notably, a BF_{10}= 245 suggests that the observed data are 245 more likely under H_1 than under H_0 (Fig. 43). The median of the resulting posterior distribution for δ equals -1.204 with a central 95% credible interval for δ that ranges from -1.889 to -0.561.

MOCA_T

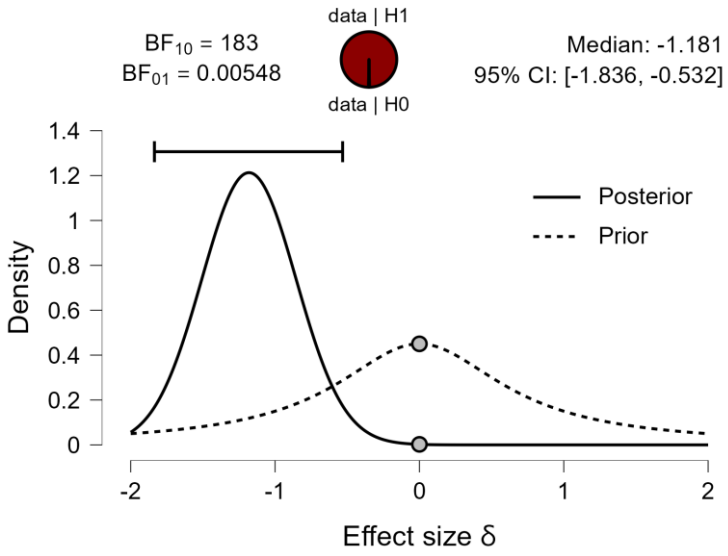


Figure 42. Prior and Posterior plot reports the Bayesian Mann-Whitney U Test results for the parameter δ (MOCA_T). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The posterior distribution's median and 95% central credible interval is displayed in the top right corner. Image created with JASP.

MS360_T

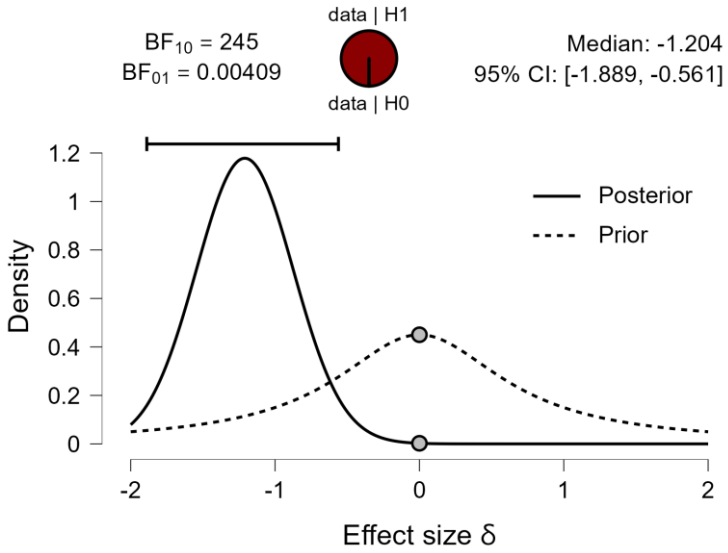


Figure 43. Prior and Posterior plot reports the Bayesian Mann-Whitney U Test results for the parameter δ (MS360_T). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The posterior distribution's median and 95% central credible interval is displayed in the top right corner. Image created with JASP.

Additionally, a One-Way Analysis of Covariance (ANCOVA) was conducted to determine if the significant difference in MS-360° scores between groups could be attributed to the Age difference between them, with MS360_T as the dependent variable, Group as the fixed factor, and Age as the covariate. JASP priors (r scale prior width = 0.5 for fixed effects, r scale prior width = 0.354 for covariates) were used. The results showed that the observed data appeared more likely under the model which included both Age and Group as predictors (Table 10), suggesting that age alone is not enough to explain the observed data.

Model Comparison ▼

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Group + Age	0.250	0.985	203.454	1.000	
Group	0.250	0.012	0.038	0.013	1.124
Age	0.250	0.002	0.006	0.002	1.124
Null model	0.250	9.699×10^{-9}	2.910×10^{-8}	9.842×10^{-9}	1.124

Table 10. Comparison of models based on the MS360_T score in the two groups. The table includes columns for the predictors used in each model (Models), the prior P(M) and posterior P(M|data) probabilities of the model (BF_M), the posterior odds of the model, and the Bayes factors for each model compared to the best model (BF_{10}). The final column estimates the numerical error in calculating the Bayes factor. The models are listed in order of lowest to highest Bayes factor about the best model.

Passing to Correlations between MOCA and MS-360° total scores, for HC group results showed moderate evidence supporting the null hypothesis, suggesting the absence of a correlation between the two scores in HC. In particular, a BF_{01}=4.182 indicates that the observed data are 4.182 times more likely under H_0 (Fig. 44).

HC Group

MOCA_T – MS360_T

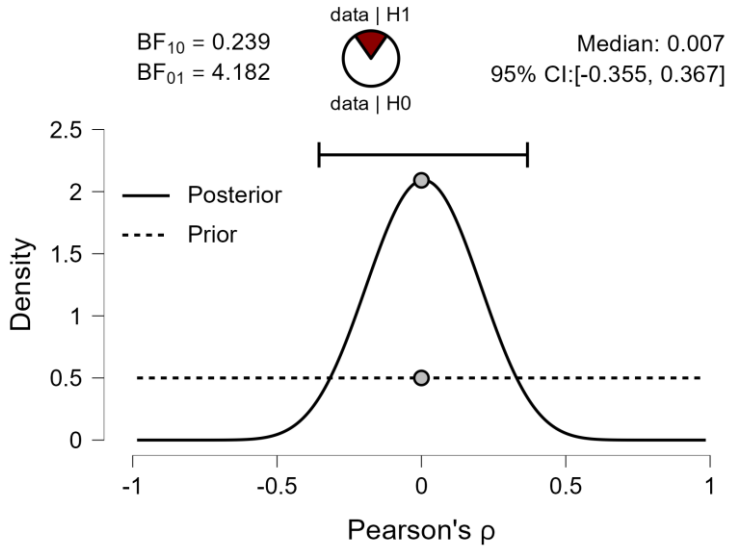


Figure 44. Prior and Posterior plot report the correlation test results for the parameter ρ (MOCA_T – MS360_T, HC Group). The reported analysis concerned the correlation between MOCA_T and MS360_T. The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The posterior distribution's median and 95% central credible interval is displayed in the top right corner. Abbreviations – MOCA_T: MOCA total score; MS360_T: MS-360° total score. Image created with JASP.

On the contrary, there is an extreme amount of evidence for the SCI group suggesting a positive correlation between the two measures. In particular, a BF_{10}=640 indicates that the observed data are 640 times more likely under H_1 (Fig. 45). The median of the resulting posterior distribution for ρ equals 0.735 with a central 95% credible interval for ρ that ranges from 0.472 to 0.901. The scatterplot presenting this correlation is shown in Figure 46.

SCI Group

MOCA_T – MS360_T

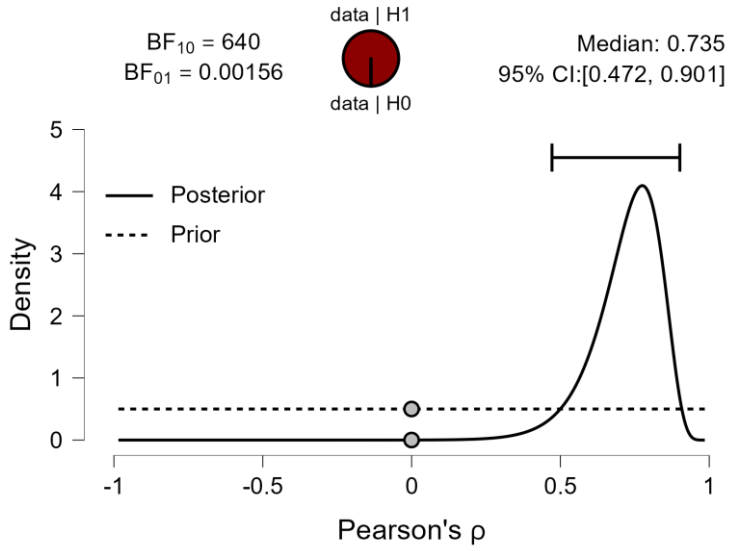


Figure 45. Prior and Posterior plot report the correlation test results for the parameter ρ (MOCA_T – MS360_T, SCI Group). The reported analysis concerned the correlation between MOCA_T and MS360_T. The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The posterior distribution's median and 95% central credible interval is displayed in the top right corner. Abbreviations - MOCA_T: MOCA total score; MS360_T: MS-360° total score. Image created with JASP.

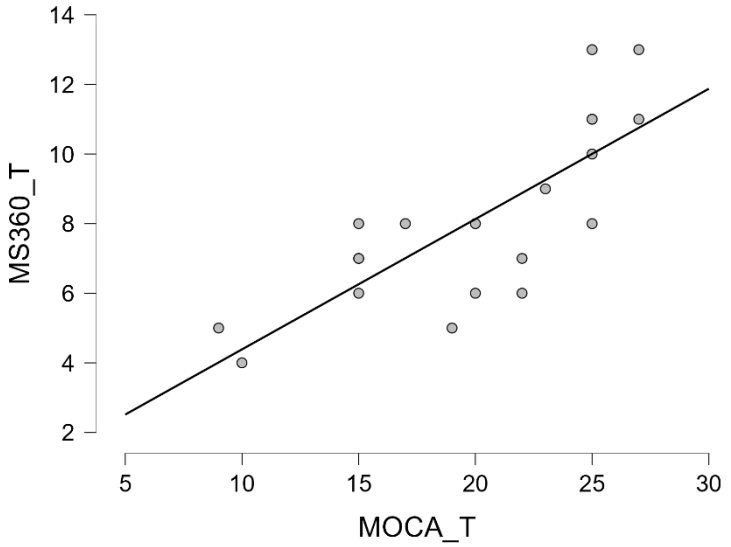
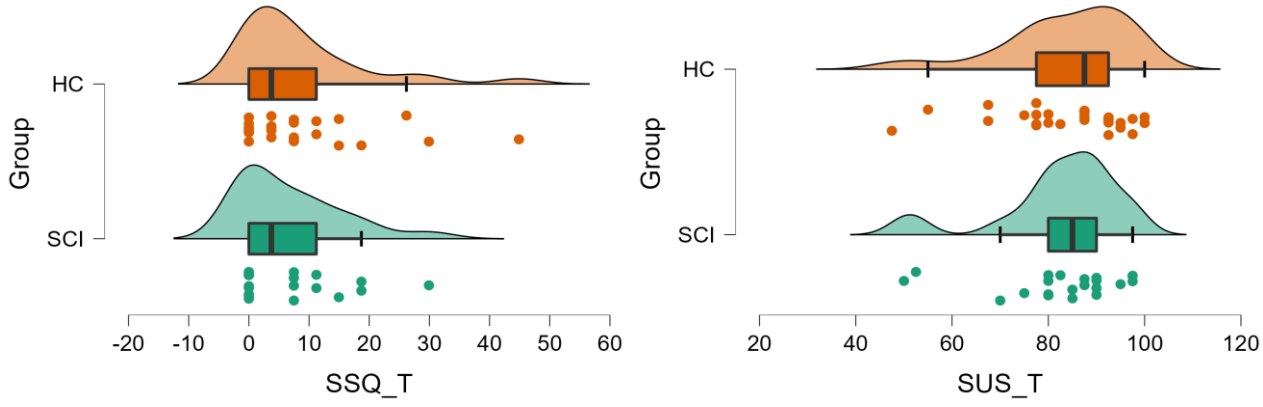


Figure 46. The scatterplot shows the SCI group's correlation between the MOCA (MOCA_T) and MS-360° (MS360_T) total scores. A strong positive correlation (r=-0.78) suggests a convergence between the two tests.

4.3.3 User-Experience

Here the dependent variables of interest were the SSQ and SUS total scores given by participants after the exposition to the MS-360°. The descriptives data concerning these two variables are presented in Figure 47.



Descriptive Statistics

	SSQ_T		SUS_T	
	SCI	HC	SCI	HC
Valid	20	27	20	27
Missing	0	0	0	0
Mean	6.732	8.588	82.250	83.611
Std. Deviation	8.546	10.771	12.745	13.107
Minimum	0.000	0.000	50.000	47.500
Maximum	29.920	44.880	97.500	100.000

Figure 47. Raincloud plots show SSQ (SSQ_T) and SUS (SUS_T) total scores distribution in the two groups: Healthy Controls (HC) and Subjective Cognitive Impairment (SCI). The descriptives table is reported under the two plots. Images created with JASP.

Starting with SSQ total score (SSQ_T), the HC group reported a mean score of 6.73 (SD=8.5, range= 0-29.9), while the SCI group reported a mean score of 8.54 (SD=10.77, range = 0-44.8). For both groups, the reported scores were positioned widely below the 95th percentile value of 44.9, indicated by the data presented in the SSQ original paper (Kennedy et al, 1993).

For what concerned perceived usability, the HC group reported a mean SUS total score (SUS_T) of 83.6 (SD= 13.1, range= 47.5-100), while the SCI group reported a mean SUS total score of 82.25 (SD= 12.74, range= 50-97.5). Both SUS_T scores fall within the "Excellent" label defined by the adjective rating scale proposed by Bangor et al. (2009) (Fig. 48).

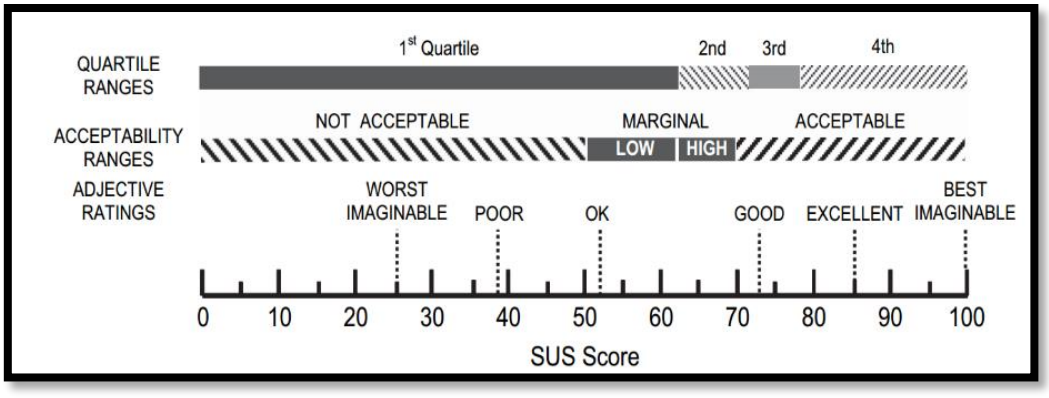


Figure 48. The adjective rating scale for SUS scores interpretation proposed by Bangor et al. (2009). The 83.6 and 82.25 mean scores given by HC and SCI participants, respectively, marks the perceived usability of the MS-360° as "Excellent".

Passing to group comparisons, to evaluate the statistical significance of the differences detected in the SSQ and the SUS total scores, two Bayesian Mann-Whitney U Tests were performed.

Starting with SSQ_T scores, the resulting BF_{01}= 2.57 showed anecdotal evidence supporting H_0 over H_1 (Fig. 49). A similar result was detected for the SUS_T scores, with a resulting BF_{01}=3.12 which suggests moderate evidence supporting H_0 over H_1 (Fig. 50).

SSQ_T

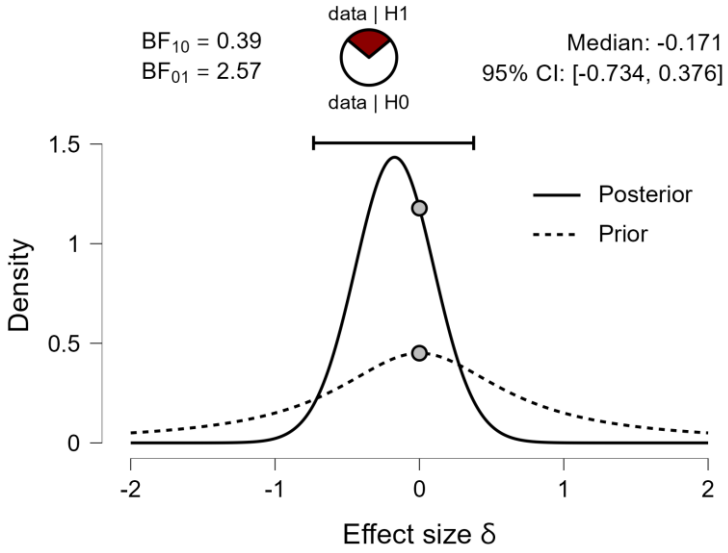


Figure 139. Prior and Posterior plot reports the Bayesian Mann-Whitney U Test results for the parameter δ (SSQ_T). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The posterior distribution's median and 95% central credible interval is displayed in the top right corner. Image created with JASP.

SUS_T

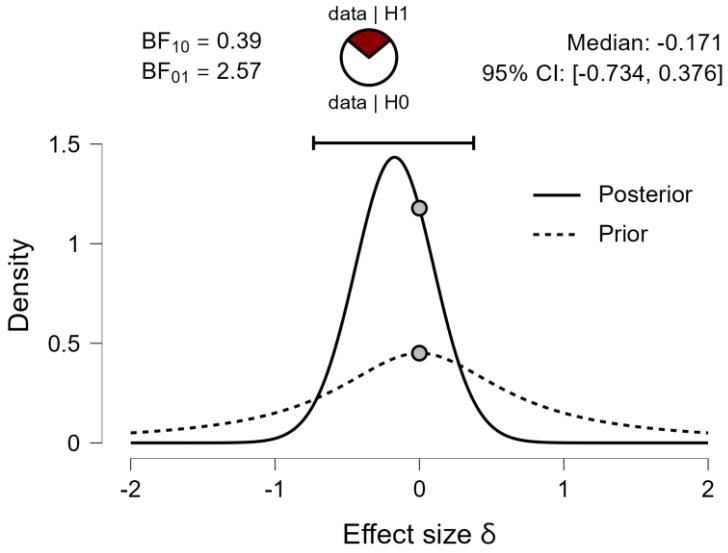


Figure 50. Prior and Posterior plot reports the Bayesian Mann-Whitney U Test results for the parameter δ (SUS_T). The top probability wheel illustrates the data's evidence under each hypothesis. The two grey dots indicate the density of the prior and posterior at the test value (Dickey and Lientz, 1970). The posterior distribution's median and 95% central credible interval is displayed in the top right corner. Image created with JASP.

4.4 Discussion

The rationale of the present study relied on evidence suggesting that 360°-VR environments can provide a more ecological delivery of neuropsychological tasks compared to Model-Based VR ones, as shown in Chapter 3. In addition to this fundamental feature, 360°-VR neuropsychological tests have been shown to be valid and sufficiently sensible to detect cognitive impairment, resulting usable to users (Serino et al., 2017; Realdon et al., 2019; Pieri et al., 2021; Bruni et al., 2022). However, until now, all the developed tools targeted only a single cognitive function and proposed single tasks. The study presented here wanted to expand the current framework concerning the development of a 360°-VR neuropsychological test, offering a pilot tool providing a first level multidomain assessment of cognitive functioning. To deliver a first glance at how a more complex tool such as the described MS-360° could perform in the field of cognitive assessment, a gold-standard paper-and-pencil tool, the MOCA, was chosen for direct comparison. Both devices were administered to an HC group and a group reporting SCI.

First, for what concerned Test Performance, the analyses were performed considering the total scores obtained by the two groups in the MOCA and the MS-360°. Results showed that the MS-360° distinguishes the average SCI performances from the HC controls, as well as MOCA. Moreover, a strong positive correlation (r=0.78) between the two test scores was detected within the SCI group, supporting convergent validity for the MS-360°. The same correlation was not found for the HC group, suggesting that the observed relationship was specific to the condition reported by SCI participants. Even if preliminary from a psychometric point of view, these results showed that the MS-360° could be a valid and sensible tool to detect possible cognitive impairment.

Moreover, the results from the ANCOVA specified that the difference between the HC and SCI groups could not be completely accounted for by the slightly different ages of the two groups.

Additionally, the results concerning simulation sickness assessment showed that, on average both groups did not suffer any adverse symptoms after MS-360° administration, with no group reporting more sickness than the another. This is an expected result, as the tool was built following the good practices for designing immersive VR applications for older adults (for a review, see Ijaz et al., 2022). Likewise, there was no difference in the perceived usability between groups, suggesting that the SCI group, although showing a worse

cognitive performance, did not perceive the MS-360° as less usable. Indeed, both groups considered the MS-360° as very usable, confirming that the greater simplicity of immersive 360°-VR interaction design is suitable for older end-users, even with suspected cognitive impairment.

However, as a pilot tool, the MS-360° has several limitations related to technology and the methodology used in the current study. First, this version did not exploit the full advantage offered by VR technology as a computerized paradigm (see Paragraph 2.1). In fact, the collection of the users' responses and the subsequent computation of the scores was done manually. Moreover, the detected performance measures did not differ from the ones usually collected in classic neuropsychological tests. The next Chapter will present a pilot study aimed at overcoming these technical limitations, providing insights into advanced data collection methods in 360°-VR environments.

The recruited samples presented a low number of participants, which limited the generalizability of the results. The small sample size also impacted the adequacy of the defined cut-off used for computing the MS-360° total score. Indeed, clues from healthy controls performance were taken to set reasonable cut-offs; a complete standardization will need to collect data from a sizeable, stratified population sample to develop more precise cut-offs to the subscores. Nonetheless, it should be appreciated that despite the current scoring being theoretically based, the convergence toward the MOCA in the SCI groups suggests that the score is meaningful, and thus the cut-offs set reasonable.

Following the considerations presented in Chapter 2, once a more advanced version of the MS-360° will be developed, the next steps require planning a proper validation study and, more crucially, the collection of normative data.

5. Exploring the Use of Advanced Data Collection in 360°-VR Settings

5.1 Theoretical and Methodological Background

In Chapter 2 it was showed how VR technology can provide a unique solution for collecting advanced data linked to different aspects of cognitive functioning. Indeed, as a computerised paradigm, it allows automatic data collection and an error-free computation of performance-related scores. Moreover, recent HMDs integrate sensors and software solutions which can be used both to collect insightful type of data and to register user's responses in a more systematic way.

As previously discussed (Chapter 3), 360°-VR scenarios allow a more ecological delivery of task-related stimuli, being more limited for what concern the kind of interaction paradigm to implement. Additionally, the objects included in these scenarios are not 3D entities which can have transform, scale, or position proprieties. They are indeed part of a photograph or a video: thus, their behaviour cannot be programmed or registered with the same level of complexity offered by Model-Based VR. This limits the choice concerning what kind of measures can be collected. The challenge is to identify first the type of behavioural responses of interest, and then the most suitable paradigm to collect them considering the abovementioned limitations which characterise 360°-VR.

The first version of the MS-360° presented in Chapter 4 did not exploit the advantages offered by the adoption of the VR paradigm, as it was more focused on defining which task to implement using 360°-VR. Thus, the present work must be considered a pilot study to test the feasibility of introducing advanced data collection within the 360°-VR environments used in the MS-360°.

5.2 Methods

5.2.1 Application Design and Task Administration

The VEs which were chosen for further development were the Target Selection Task (TS) (see Paragraph 3.2.2.2) and the Free Visual Exploration Task (FVE) (see Paragraph 3.3.1.2). A new HMD was adopted for this study: the Pico Neo Pro Eye 3 (Fig. 51). This device was selected because it

offers the same characteristics as the Meta Quest 2 but also integrates an advantageous eye-tracker sensor. Moreover, the resulting experience is not PC-tethered anymore but can run stand-alone.

Given their characteristics (i.e., visual exploration and object naming), these tasks were chosen as suitable candidates for fully exploiting the possibilities offered by having a dedicated eye-tracking device and a speech-to-text algorithm with keyword recognition capabilities. The two fundamental paradigms are the ones which will then be extended to the other tasks of the MS-360°.



Figure 514. A photograph of the headset used in the study, the Pico Neo Pro 3 Eye (https://www.picoxr.com/it/products/neo3-pro-eye). This device supports both 3 and 6-DoF standalone experiences. Additionally, it is equipped with an eye-tracking sensor. The picture was taken by me at MiBTec (https://www.mibtec.it/en/), University of Milano-Bicocca, Milan, Italy.

The experience consisted of a familiarisation phase and an assessment phase.

The familiarisation phase is not considered in calculating the final test score. Still, it is part of the experience as it instructs the users on the exploration modality and the use of controllers and, eventually for detecting any side effects related to VR exposure (e.g., cybersickness). Starting the experience, users find themselves in a neutral environment where they are given auditory instructions.

They can choose whether to listen again or to start the task by selecting one of the two buttons on the user interface (Fig. 52a). The instructions invite them to explore the environment and select different buttons. Users must trigger a lever on the controller to interact with the scenario's elements. Throughout the experience, they are seated on a swivel chair to explore the 3-DoF VE more smoothly. When this phase is completed, they are invited to proceed to the assessment phase.

The assessment phase consists of two tasks. The first is the FVE. Users are placed at the centre of a living room, and when the task starts, they are required to explore the environment and name all the objects they identify within a minute. Eye-tracking technology makes this task more sensitive (Kaufmann et al., 2020). For these reasons, two background processes have been added to the system that runs in parallel while the user is speaking. The first consists of recognizing the pronounced words, transcribing and labelling them: using an external service to translate words in the text (Microsoft Azure), the system checks their correctness and eventually registers the timestamp. The second one, exploiting the embedded eye-tracking sensor in the HMD, creates a mapping of the visual exploration, producing a heatmap. The second task is the TS. Also, in this setting, users are in the centre of the same living room, but yellow and green cups and glasses have been added to the scene. When the task starts, users must scan for the yellow cups and use the controller raycaster to point and select them (Fig. 52b).



Figure 52. (a) An example of the user-interface interaction within the application. (b) A screenshot of a user executing the Target Selection Task.

The system computes the final score of the task relying on the concept of the Signal Detection Theory (SDT) (Macmillan, 2002). The starting point for SDT is that nearly all reasoning and decision-making occur in some uncertainty (Brucki and Nitrini, 2008). To recreate this condition, 36 stimuli were used in the scene. Of these, only 11 are defined as targets (yellow glasses), while the remaining 25 (green cups and glasses) are distractors. The user must be able to recognise the target objects and select only the yellow glasses. To translate the users' actions into quantitative data, the same metrics found in the SDT were used, namely:

• Hit Rate (Hit): the number of target trials selected over the total number of them.

• False Alarm Rate (FA): the number of non-target trials selected over the total number of green glasses.

• Sensitivity (d'): the difference between the z-transforms of the Hit Rate and the False Alarm Rate. It represents the final score of the task.

5.2.2 Experimental Procedure

To adopt automatic scoring in clinical practice, the reliability of the data collection must be verified. Therefore, before extending the eye tracking and the speech-to-text procedure to the entire MS-360 protocol and conducting an extensive validation study on healthy and clinical populations, the quality of the automatic scoring needed to be proven. The following study is an explorative investigation of the reliability of automated scoring and the system's usability. The study was conducted on eleven users to test the features implemented in the system, focusing on the proper functioning of the system itself and the collection of data.

The study involved N=11 healthy young-adult participants (aged 21 to 26). Each subject had little or no experience with VR. All participants were contacted in advance and were recruited voluntarily. Exclusion criteria, study procedure, goals, and data treatment were provided through appropriate documentation. Before the beginning of the study, informed consent was collected. Furthermore, each participant could leave the study at any time. The study was conducted in collaboration with the Politecnico of Milan. The Ethical Committees of Politecnico of Milan approved the study protocol and authorised its execution. The conditions and restrictions dictated by Covid19 negatively impacted the recruitment phase. For this reason, it was decided to recruit a sample of healthy subjects rather than the ideal target of the application. Although this is sub-optimal to our study, the usability and acceptability of VR in elderly patients have already been established in several studies (Huygelier et al., 2019; Tuena et al., 2020), which did not report disruptive problems in using VR technology in elders.

The protocol consisted of three phases with a total session duration of about 20 minutes. The first phase served as an introduction to explain the project's context and start the tutorial of the experience (familiarisation phase). In the second phase, the user started the actual test. During the assessment, the participant performed both tasks, the FVE task and the TS task. Participants sat on a turning chair, allowing a 3-DoF exploration of the VEs. A team member was close to the participant throughout the

study to help if necessary. The last phase was dedicated to the user completing the SUS (Brooke, 1996). All answers were collected with a Google form.

Data gathering methods then included:

• Quantitative measures:

–FVE: number of keywords correctly recognised by the system, timestamp for each pronounced word.

– TS: timestamp for each selection (target and non-target), Hit, FA, and d' scores.

– Usability: SUS scores.

• Qualitative measures as direct observations of users' behaviours and comments during the performances.

Gathering all necessary information, appropriate behaviours and comments was managed by manually reporting observations on a standardised form. Being a preliminary study, the testing focused on obtaining results that could be evaluated and collected quickly. Then SUS was chosen considering the limited number of items compared to other estimating methods and the evidence supporting its use in assessing VR systems (Pedroli et al., 2015; Dulau et al., 2019; Tuena et al., 2020; Pieri et al., 2021).

5.3 Results

5.3.1 Keyword Recognition

Figure 53 displays the aggregated data for keyword recognition. As the number of pronounced keywords increases, the error rate remains low. It should be noted, however, that the sample size for this study is small, and caution should be exercised when interpreting these data.

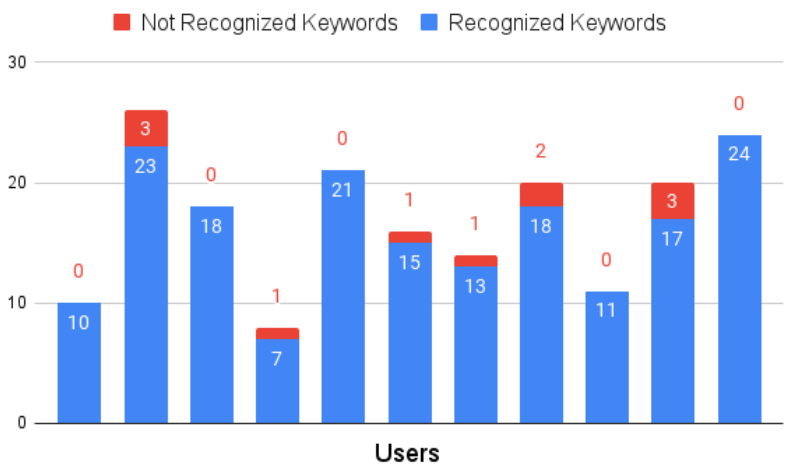


Figure 53. A histogram shows each user recognised and not recognised keywords.

5.3.2 Heatmap

During the execution of the FVE task, the system captured the visual exploration of the participant, producing a heatmap for each of them (Fig. 54).

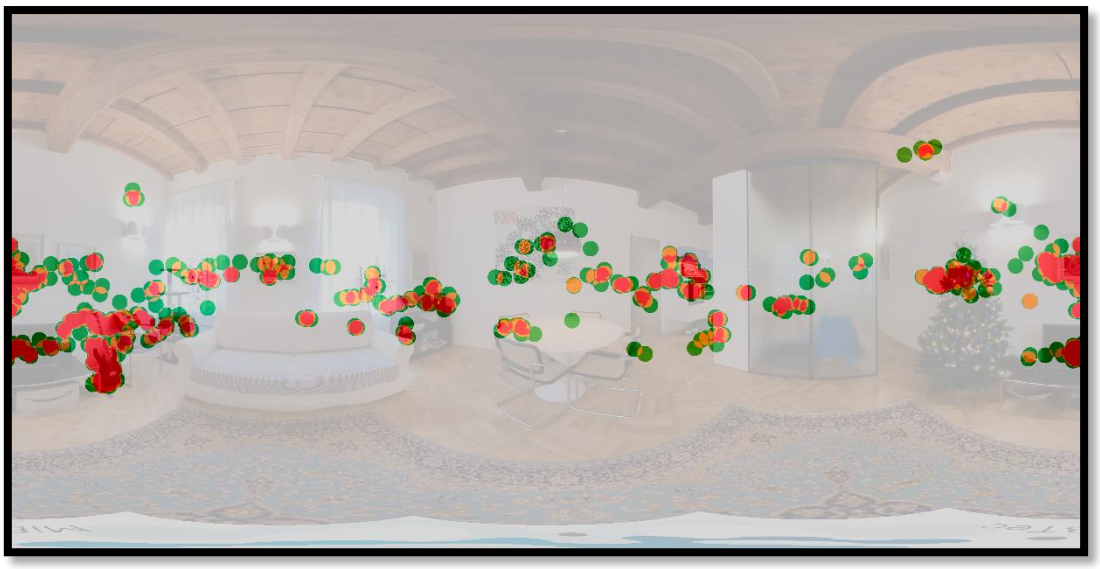


Figure 54. Heatmap showing the performance of one user on the FVE task, superimposed on the equirectangular projection of the 360°-VR environment. The heatmap is colour-coded, with green representing a low number of fixations and red representing a high number of fixations.

From a qualitative perspective, it was observed that the system correctly associated each named object with the explorative pattern showed by participants.

5.3.3 SUS Questionnaire

The SUS total median value (Fig. 55) is 85 (IQR=15), with a minimum rating score of 62.5 and a maximum rating score of 95. The mean score of 83,6 (SD=10,2) shows that the system is perceived as "good" and "acceptable" according to the adjective scale developed by Bangor et al. (2009). SUS results related to the single items are represented in Figure 56. The higher scores were given to things i3 ("I thought the app was easy to use") and i7 ("I would imagine that most people would learn to use this app very quickly"), both concerning the ease of use, with 91% of respondents reporting a score equal or higher than 4 out of five. Noteworthy are the low scores from i6 ("I found inconsistency between the various functionalities of the system").

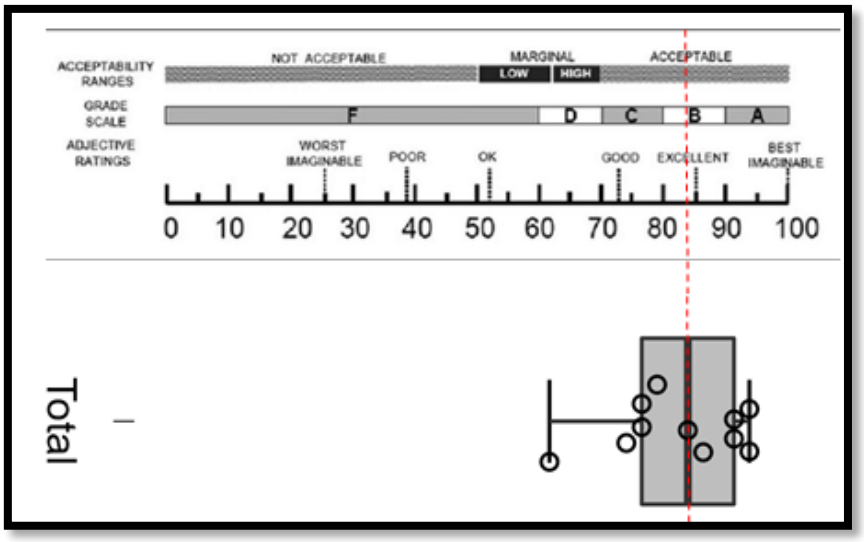


Figure 55. A boxplot of the SUS scores embedded within the adjective scale was developed by Bangor et al., (2009).

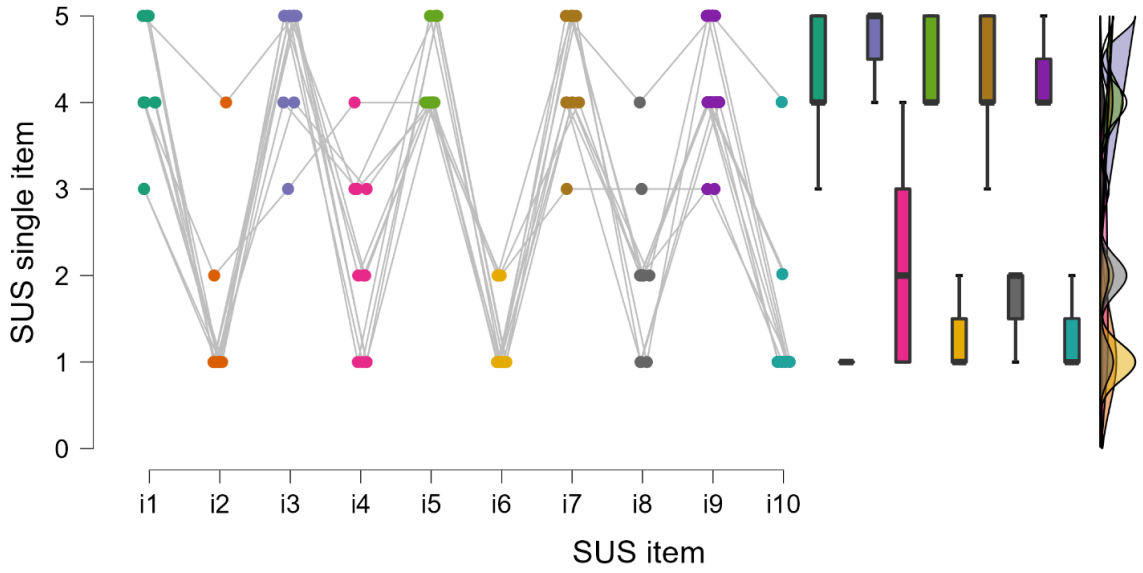


Figure 56. A graph showing the distribution of scores given by the participants for each item on SUS is represented as a raincloud plot.

5.4 Discussion

Results show that the external services used for the system's scripts are accurate and reliable. The focus was on ensuring that the system was capturing data continuously without bugs or errors rather than on testing the external services or the cognitive meaning of the results. SUS results suggest that the user's interaction with the system is consistent throughout the experience. While these results are preliminary, especially considering that the sample participating in the study was not the target audience for the application, they are promising also considering that participants were utterly naïve to using VR.

Observation at a qualitative level indicated that exploration mapping aligned with the automatically detected naming list, suggesting that both were functioning correctly. Without considering the person's actual skills, it can be affirmed that when the user's eyes pointed to a specific coordinate in the VE, the custom script correctly identified the corresponding pixel in the output image as a fixation point. This suggests that the system performed well.

Since the system automatically calculates the Hit, FA, and d' scores from the target selection task, further analysis was optional. However, it was noted that users were trying multiple times to select the same glass because there was no feedback after the "select" action. To address this issue, non-intrusive haptic feedback was added to the system whenever the trigger was pressed during the scene's entire duration to provide the user confidence while completing the task.

The purpose of this exploratory study was to test the usability of a system integrating 360°VEs with an automated data collection tool for the assessment of cognitive functions based on the ongoing debate regarding the ecological validity of classical tests and on the evidence that supports the use of VR in neuropsychological assessment (Parsons, 2015). The results, although preliminary, are encouraging and show that the technology above can be used more extensively, and it is promising to provide more details as an assessment tool in neuropsychological tests. Today's technology is ready to be fully exploited, and tomorrow's HMD will be increasingly powerful and precise to provide more accurate measurements.

Next Steps: Once the system's usability has been ascertained, the following steps will evaluate its effectiveness as a neurocognitive assessment tool. First, the automatic scoring needs to be extended to all the subtests of the MS-360°. Second, it will be necessary to administer the test to a broader sample of users to validate the tool and obtain normative data. Third, it will require a selection close to the intended target group, i.e., people with suspected cognitive impairment. This would also allow us to gather additional feedback on the system's usability from older adults and people with initial or suspected cognitive problems.

6. Conclusion and Future Perspectives

This three-year-long work aimed to investigate the possible benefits and challenges derived from the implementation of 360°-VR scenarios in developing novel neuropsychological tests. The results from the systematic review presented in Chapter 2 showed that VR had reached a level of methodological maturity making it a possible candidate for bringing a paradigm shift in the field of neuropsychological assessment. However, currently (2023), only three VR-based tests were found to be supported by all the psychometric evidence needed to demonstrate their suitability for everyday clinical practice. Moreover, most new VR-based tools were developed using Model-Based VR, with only a few studies investigating the possible use of 360°-VR. The present work first aimed at testing the theoretical hypothesis that the photorealism provided by 360-°VR could further improve the ecological validity of the assessment procedures when compared to Model-Based VR. The two experiments presented in Chapter 3 showed that 360°-VR is perceived as more ecological regarding the given environment's realism. However, from a technological perspective, it was shown and discussed how 360°-VR can provide only limited interaction with the environment. Thus, this suggests that 360°-VR is not more ecological in absolute terms but only for what concerns stimulus delivery. On the other hand, Model-Based VR may be more ecological about the verisimilitude of the responses required by the users. The choice regarding which type of VR environment to use to implement a novel neuropsychological test may be determined by the kind of task to simulate. For example, it is possible to imagine that tasks aimed at testing praxis functions are more suitable for Model-Based scenarios. At the same time, memory tests can benefit more from the realism provided by 360°-VR. Indeed, the firsts usually require users to manipulate objects, which is allowed only by Model-Based VR. On the other hand, testing memory encoding processes using photorealistic stimuli can provide more meaningful measures of everyday memory functioning (Tatler and Melcher, 2007).

However, the simpler interaction design which characterises 360°-VR is not only a limitation, as it can guarantee a good level of accessibility and usability. For example, the pilot version of the MS-360° presented in Chapter 4 included a straightforward 3-DoF interaction design based on a

raycasting selection of buttons. Even with different levels of cognitive impairment, all users were very engaged during the execution of each task. At the end of the experiment, they were impressed by the realism of each presented scene. Moreover, on average, participants reported a high level of perceived usability. These results, added to those showing the sensibility of MS-360° in detecting cognitive impairment, supported a further extension of this pilot neuropsychological test. Consequently, Chapter 5 described an example of advanced data collection methods which can be integrated into the future developments of MS-360°. Considering the abovementioned limitations concerning the kind of users' response to collect within 360°-VR environments, two paradigms were selected to be more suitable for detecting and computing users' responses: eye-tracking and speech recognition. These modalities were chosen because they can provide full access to the only two functional channels not limited by 360°-VR design impositions: eye behaviour and speech. The results from the pilot study showed that the integration of these advanced methods was successful for two tasks of the MS-360°, supporting the extension of this paradigm to the other tasks of the test.

In conclusion, the presented work provided a well-focused step toward developing a ready-to-use 360°-VR neuropsychological test. The COVID-19 pandemic heavily downsized the original objectives which were set for this project, which were obtaining normative data for the screening test. In Italy the access to research laboratories was very much limited for about the first year and a half of this project's lifespan. This specifically impacted on the availability and numerosity of recruitable participants to reach the numerosity of a normative sample for the MS-360, and the participation of fragile individuals with cognitive impairment. The original project objectives become the future steps needed to continue with the development of the MS-360°, which integrate with the more advanced interaction design developed in the meanwhile. The author intends to continue the validation process by recruiting more numerous and heterogeneous samples of participants, being healthy older adults providing normative data, or cognitive impaired individuals providing validity and sensitivity data.

The same must be done for testing the usability of a more technologically advanced version of the MS-360° characterised by a revised interaction design.

Reference List

1. Acer (2019) *Windows Mixed Reality Headset Design* [Infographic]. Acer United Kingdom. https://www.acer.com/ac/en/GB/content/series-design/wmr
2. Albani, G., Pignatti, R., Bertella, L., Priano, L., Semenza C, Molinari, E., Riva, G., and Mauro, A. (2002). Common daily activities in the virtual environment: a preliminary study in parkinsonian patients. *Neurological Sciences*, *23*(2), 49–50. https://doi.org/10.1007/s100720200064
3. Allain, P., Foloppe, D. A., Besnard, J., Yamaguchi, T., Etcharry-Bouyx, F., Le Gall, D., Nolin, P., and Richard, P. (2014). Detecting Everyday Action Deficits in Alzheimer's Disease Using a Nonimmersive Virtual Reality Kitchen. *Journal Of the International Neuropsychological Society*, *20*(5), 468–477. https://doi.org/10.1017/S1355617714000344
4. Areces, D., Dockrell, J., García, T., Gonzaález-Castro, P., and Rodríguez, C. (2018a). Analysis of cognitive and attentional profiles in children with and without ADHD using an innovative virtual reality tool. *PLoS ONE*, *13*(8). https://doi.org/10.1371/journal.pone.0201039
5. Areces, D., Rodríguez, C., García, T., Cueli, M., and González-Castro, P. (2018b). Efficacy of a Continuous Performance Test Based on Virtual Reality in the Diagnosis of ADHD and Its Clinical Presentations. *Journal of Attention Disorders*, *22*(11), 1081–1091. https://doi.org/10.1177/1087054716629711
6. Atkins, A. S., Khan, A., Ulshen, D., Vaughan, A., Balentin, D., Dickerson, H., Liharska, L. E., Plassman, B., Welsh-Bohmer, K., and Keefe, R. S. E. (2018). Assessment of Instrumental Activities of Daily Living in Older Adults with Subjective Cognitive Decline Using the Virtual Reality Functional Capacity Assessment Tool (VRFCAT). *The Journal of Prevention of Alzheimer's Disease*, *5*(4), 216–234. https://doi.org/10.14283/jpad.2018.28

7. Bangor A, Philip TK, James TM (2008) An empirical evaluation of the system usability scale. *Int J Hum Comput Int* 24:574–594. https://doi.org/10.1080/10447310802205776
8. Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved.* The Guilford Press.
9. Barnard D. (2019). *Image showing the difference between 3-DoF (rotational movement) and 6-DoF (rotational and translational movement) with a VR headset* [Infographic]. Virtualspeech Blog. https://vs-static.virtualspeech.com/img/blog/3dof-6dof-vr-headset.jpg
10. Bechara, A., Damasio, A. R., Damasio, H., and Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50(1-3),* 7-15.
11. Bell, I. H., Nicholas, J., Alvarez-Jimenez, M., Thompson, A., and Valmaggia, L. (2020). Virtual reality as a clinical tool in mental health research and practice. *Dialogues in clinical neuroscience*, *22*(2), 169–177. https://doi.org/10.31887/DCNS.2020.22.2/lvalmaggia
12. Belleville, S., Rouleau, N., Van der Linden, M., and Collette, F. (2003). Effect of manipulation and irrelevant noise on working memory capacity of patients with Alzheimer's dementia. *Neuropsychology, 17(1),* 69–81. https://doi.org/10.1037/0894-4105.17.1.69
13. Bergouignan, L., Nyberg, L., and Ehrsson, H. H. (2014). Out-of-body-induced hippocampal amnesia. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(12), 4421–4426. https://doi.org/10.1073/pnas.1318801111
14. Besnard, J., Richard, P., Banville, F., Nolin, P., Aubin, G., Le Gall, D., Richard, I., and Allain, P. (2016). Virtual reality and neuropsychological assessment: The reliability of a virtual kitchen to assess daily-life activities in victims of traumatic brain injury. *Applied Neuropsychology: Adult*, *23*(3), 223–235. https://doi.org/10.1080/23279095.2015.1048514
15. Bilder, R. M. (2011). Neuropsychology 3.0: Evidence-Based science and practice. *Journal of the International Neuropsychological Society, 17(1),* 7–13. https://doi.org/10.1017/S1355617710001396

16. Bioulac, S., Lallemand, S., Rizzo, A., Philip, P., Fabrigoule, C., and Bouvard, M. P. (2012). Impact of time on task on ADHD patient's performances in a virtual classroom. *European Journal of Paediatric Neurology: EJPN: Official Journal of the European Paediatric Neurology Society*, *16*(5), 514–521. https://doi.org/10.1016/j.ejpn.2012.01.006
17. Brooke J (1996) SUS: a "quick and dirty" usability scale. In: Jordan P, Thomas W, Weerdmeester A, McClelland I (eds) Usability evaluation in industry. Taylor and Francis, London, pp 189–194
18. Brucki, S. M., and Nitrini, R. (2008). Cancellation task in very low educated people. *Archives of clinical neuropsychology: the official journal of the National Academy of Neuropsychologists*, 23(2), 139–147. https://doi.org/10.1016/j.acn.2007.11.003
19. Bruni F, Mancuso V, Stramba-Badiale C, Greci L, Cavallo M, Borghesi F, Riva G, Cipresso P, Stramba-Badiale M and Pedroli E (2022) ObReco-2: Two-step validation of a tool to assess memory deficits using 360° videos. *Front. Aging Neurosci*. 14:875748. doi: 10.3389/fnagi.2022.875748
20. Buckingham G (2021) Hand Tracking for Immersive Virtual Reality: Opportunities and Challenges. *Front. Virtual Real.* 2:728461. doi: 10.3389/frvir.2021.728461
21. Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L. M. A., Dawson, D. R., Anderson, N. D., Gilbert, S. J., Dumontheil, I., and Channon, S. (2006). The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society*, *12*(2), 194–209. https://doi.org/10.1017/S1355617706060310
22. Buxbaum, L. J., Palermo, M. A., Mastrogiovanni, D., Read, M. S., Rosenberg-Pitonyak, E., Rizzo, A. A., and Coslett, H. B. (2008). Assessment of spatial attention and neglect with a virtual wheelchair navigation task. *Journal of Clinical and Experimental Neuropsychology*, *30*(6), 650–660. https://doi.org/10.1080/13803390701625821

23. Buxbaum, L. J., Dawson, A. M., and Linsley, D. (2012). Reliability and validity of the virtual reality lateralized attention test in assessing hemispatial neglect in right-hemisphere stroke. *Neuropsychology, 26(4),* 430–441. https://doi.org/10.1037/a0028674

24. Buytaert, D. (2017). Shopping with Augmented Reality. [Photograph]. Dries Buytaert Blog. https://dri.es/files/images/blog/shopping-with-augmented-reality-example-2017.jpg

25. Cabinio, M., Rossetto, F., Isernia, S., Saibene, F. L., Di Cesare, M., Borgnis, F., Pazzi, S., Migliazza, T., Alberoni, M., Blasi, V., and Baglio, F. (2020). The Use of a Virtual Reality Platform for the Assessment of the Memory Decline and the Hippocampal Neural Injury in Subjects with Mild Cognitive Impairment: The Validity of Smart Aging Serious Game (SASG). *Journal of clinical medicine*, *9*(5), 1355. https://doi.org/10.3390/jcm9051355

26. Camina, E., and Güell, F. (2017). The neuroanatomical, neurophysiological and psychological basis of memory: Current models and their origins. *Frontiers in Pharmacology, 8(JUN),* 438. https://doi.org/10.3389/fphar.2017.00438

27. Cánovas, R., León, I., Roldán, M. D., Astur, R., and Cimadevilla, J. M. (2009). Virtual reality tasks disclose spatial memory alterations in fibromyalgia. *Rheumatology, 48(10),* 1273–1278. https://doi.org/10.1093/rheumatology/kep218

28. Canty, A. L., Fleming, J., Patterson, F., Green, H. J., Man, D., and Shum, D. H. K. (2014). Evaluation of a virtual reality prospective memory task for use with individuals with severe traumatic brain injury. *Neuropsychological Rehabilitation*, *24*(2), 238–265. https://doi.org/10.1080/09602011.2014.881746

29. Carlesimo G.A., Caltagirone C., Gainotti G. and the Group for the standardization of the Mental Detrioration Battery (1996) The Mental Deterioration Battery: normative data, diagnostic reliability and qualitative analyses of cognitive impairment. *European Journal of Neurology* 36:378–384

30. Casaletto, K. B., and Heaton, R. K. (2017). Neuropsychological assessment: Past and future. *Journal of the International Neuropsychological Society*, 23(9-10), 778–790. https://doi.org/10.1017/S1355617717001060

31. Cavedoni, S., Chirico, A., Pedroli, E., Cipresso, P., and Riva, G. (2020). Digital Biomarkers for the Early Detection of Mild Cognitive Impairment: Artificial Intelligence Meets Virtual Reality. *Frontiers in human neuroscience*, *14*, 245. https://doi.org/10.3389/fnhum.2020.00245

32. Chalmers, A., and Ferko, A. (2008). Levels of realism: from virtual reality to real virtuality. *Spring conference on Computer graphics.*

33. Chan, R. C. K., Shum, D., Toulopoulou, T., and Chen, E. Y. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology, 23(2),* 201–216. https://doi.org/10.1016/j.acn.2007.08.010

34. Chang E., Kim H- T., Yoo B. (2020): Virtual Reality Sickness: A Review of Causes and Measurements. *International Journal of Human–Computer Interaction*, DOI: 10.1080/10447318.2020.1778351

35. Chaytor, N., Schmitter-Edgecombe, M. (2003). The Ecological Validity of Neuropsychological Tests: A Review of the Literature on Everyday Cognitive Skills. *Neuropsychol Rev* 13, 181–197 https://doi.org/10.1023/B:NERV.0000009483.91468.fb

36. Cipresso, P., Giglioli, I. A. C., Raya, M. A., and Riva, G. (2018). The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology*, 9, Article 2086. https://doi.org/10.3389/fpsyg.2018.02086

37. Cipresso, P., Albani, G., Serino, S., Pedroli, E., Pallavicini, F., Mauro, A., and Riva, G. (2014). Virtual multiple errands test (VMET): a virtual reality-based tool to detect early executive functions deficit in Parkinson's disease. *Frontiers in Behavioral Neuroscience, 8.* https://doi.org/10.3389/fnbeh.2014.00405

38. Climent, G., Rodríguez, C., García, T., Areces, D., Mejías, M., Aierbe, A., Moreno, M., Cueto, E., Castellá, J., and Feli González, M. (2021). New virtual reality tool (Nesplora Aquarium) for assessing attention and working memory in adults: A normative study. *Applied Neuropsychology:Adult*, 6, 1-13. https://doi.org/10.1080/23279095.2019.1646745
39. Cohen, R. A. (2014). *The Neuropsychology of Attention*. Springer US. https://doi.org/10.1007/978-0-387-72639-7
40. Conners, C. K. (2000). *Conners' continuous performance test*. Multi-Health Systems.
41. Corriveau Lecavalier, Ouellet, N., Boller, E., Belleville, B., and Sylvie. (2020). Use of immersive virtual reality to assess episodic memory: A validation study in older adults. *Neuropsychological Rehabilitation*, *30*(3), 462–480. https://doi.org/10.1080/09602011.2018.1477684
42. Crawford, J. R., and Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *Clinical Neuropsychologist*, *12*(4), 482–486. https://doi.org/10.1076/clin.12.4.482.7241
43. Cox, A., Jensen, M., Ruttenberg, A., Szigeti, K., and Diehl, A. D. (2013). Measuring Cognitive Functions: Hurdles in the Development of the NeuroPsychological Testing Ontology. *ICBO*.
44. Davison, S. M. C., Deeprose, C., and Terbeck, S. (2018). A comparison of immersive virtual reality with traditional neuropsychological measures in the assessment of executive functions. *Acta Neuropsychiatrica*, *30*(2), 79–89. https://doi.org/10.1017/neu.2017.14
45. Dawson, A. M., Buxbaum, L. J., and Rizzo, A. A. (2008). The virtual reality lateralized attention test: Sensitivity and validity of a new clinical tool for assessing hemispatial neglect. *2008 Virtual Rehabilitation, IWVR*, 77–82. https://doi.org/10.1109/ICVR.2008.4625140
46. D'Elia, L., Satz, P., Uchiyama, C. L., and White, T. (1996). *Color trails test*. PAR Odessa, FL.

47. Delis, D. C. (2000). *California verbal learning test. Adult Version. Manual.* Psychological Corporation.

48. Denisova, A., and Cairns, P. (2015). First person vs. Third person perspective in digital games: Do player preferences affect immersion? In CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 145-148). ACM. https://doi.org/10.1145/2702123.2702256

49. Díaz-Orueta, U., Garcia-López, C., Crespo-Eguílaz, N., Sánchez-Carpintero, R., Climent, G., and Narbona, J. (2014). AULA virtual reality test as an attention measure: Convergent validity with Conners Continuous Performance Test. *Child Neuropsychology*, *20*(3), 328–342. https://doi.org/10.1080/09297049.2013.792332

50. Drovandi, Christopher and Pettitt, Anthony and Lee, Anthony. (2015). Bayesian Indirect Inference Using a Parametric Auxiliary Model. *Statistical Science*. 30. 10.1214/14-STS498.

51. Dickey, J. M., and Lientz, B. P. (1970). The Weighted Likelihood Ratio, Sharp Hypotheses about Chances, the Order of a Markov Chain. *The Annals of Mathematical Statistics*, 41(1), 214–226. http://www.jstor.org/stable/2239734

52. Dulau, Eva and Botha-Ravyse, Chrisna and Luimula, Mika and Markopoulos, Panagiotis and Markopoulos, Evangelos. (2019). A virtual reality game for cognitive impairment screening in elderly: a user perspective. 10.1109/CogInfoCom47531.2019.9089973.

53. Earnshaw RA, Gigante MA, Jones HH, (1993), Virtual Reality Systems, Academic Press

54. Eling P. (2019). History of Neuropsychological Assessment. *Frontiers of neurology and neuroscience*, 44, 164–178. https://doi.org/10.1159/000494963

55. European Space Agency (2017). *A researcher using VR headset to investigate ideas for controlling rovers on a planet* [Photograph]. Wikimedia Commons. https://upload.wikimedia.org/wikipedia/commons/thumb/2/27/MicrosoftHoloLensRoomProjection.JPG/1200px-MicrosoftHoloLensRoomProjection.JPG?20160410190434

56. Fordell, H., Bodin, K., Bucht, G., and Malm, J. (2011). A virtual reality test battery for assessment and screening of spatial neglect. *Acta Neurologica Scandinavica, 123(3),* 167–174. https://doi.org/10.1111/j.1600-0404.2010.01390.x

57. Gallagher, M., and Ferrè, E. R. (2018). Cybersickness: a Multisensory Integration Perspective. *Multisensory research*, *31*(7), 645–674. https://doi.org/10.1163/22134808-20181293

58. Gamito, P., Morais, D., Oliveira, J., Ferreira Lopes, P., Picareli, L. F., Matias, M., Correia, S., and Brito, R. (2016a). Systemic Lisbon Battery: Normative Data for Memory and Attention Assessments. *JMIR Rehabilitation and Assistive Technologies*, *3*(1), e5. https://doi.org/10.2196/rehab.4155

59. Gamito, P., Oliveira, J., Brito, R., Lopes, P., Rodelo, L., Pinto, L., and Morais, D. (2016b). Evaluation of cognitive functions through the systemic lisbon battery: Normative data. *Methods of Information in Medicine, 55(1),* 93–97. https://doi.org/10.3414/ME14-02-0021

60. Gamito, P., Oliveira, J., Alghazzawi, D., Fardoun, H., Rosa, P., Sousa, T., Maia, I., Morais, D., Lopes, P., and Brito, R. (2017). The Art Gallery Test: A Preliminary Comparison between Traditional Neuropsychological and Ecological VR-Based Tests. *Frontiers In Psychology*, *8*, 1911. https://doi.org/10.3389/fpsyg.2017.01911

61. Garcia, Silvia and Trejo, Paulina and García, Alberto and Dumas, César. (2019). Virtual Reality and Neural Networks for Exploiting Geotechnical Data: Proceedings of the 2nd GeoMEast International Congress and Exhibition on Sustainable Civil Infrastructures, Egypt 2018 – The Official International Congress of the Soil-Structure Interaction Group in Egypt (SSIGE). 10.1007/978-3-030-01941-9_2.

62. Goss-Sampson, M. A. (2020). Bayesian Inference in JASP: A Guide for Students. May 2020.

63. Harper, D. (n.d.). Etymology of virtual. Online Etymology Dictionary. Retrieved December 9, 2022, from https://www.etymonline.com/word/virtual.

64. Harvey P. D. (2012). Clinical applications of neuropsychological assessment. *Dialogues in clinical neuroscience*, *14*(1), 91–99. https://doi.org/10.31887/DCNS.2012.14.1/pharvey

65. Higuera-Trujillo, J. L., López-Tarruella Maldonado, J., Llinares Millán, C. (2017). Psychological and physiological human responses to simulated and real environments: A comparison between Photographs, 360° Panoramas, and Virtual Reality. *Applied Ergonomics*, 65, 398–409. https://doi.org/10.1016/J.APERGO.2017.05.006

66. Hougaard, B. I., Knoche, H., Jensen, J., and Evald, L. (2021). Spatial Neglect Midline Diagnostics from Virtual Reality and Eye Tracking in a Free-Viewing Environment. *Frontiers in psychology*, *12*, 742445. https://doi.org/10.3389/fpsyg.2021.742445.

67. HTC Corporation (2022). *Setting up the play area* [Infographic].Vive Cosmos Support Page. https://www.vive.com/media/filer_public/support_zip_img/hk/www/cosmos/guid-8ea8b2b2-c463-41c9-932a-62889adadc0b-web.jpg

68. Huygelier, H., Schraepen, B., van Ee, R., Vanden Abeele, V., and Gillebert, C. R. (2019). Acceptance of immersive head-mounted virtual reality in older adults. Scientific reports, 9(1), 4519. https://doi.org/10.1038/s41598-019-41200-6

69. Ijaz, K., Tran, T. T. M., Kocaballi, A. B., Calvo, R. A., Berkovsky, S., and Ahmadpour, N. (2022). Design considerations for immersive virtual reality applications for older adults: a scoping review. *Multimodal Technologies and Interaction*, *6*(7), 1-26. [60]. https://doi.org/10.3390/mti6070060

70. Insta360 (2023). *Insta360 Pro 2 Product Features* [Infographic]. Insta360 Pro 2 Product Page. https://res.insta360.com/static/assets/storage/20201009/e3c1fdda82928fd62561559c61ce52da/10.jpg

71. Iriarte, Y., Diaz-Orueta, U., Cueto, E., Irazustabarrena, P., Banterla, F., and Climent, G. (2016). AULA—Advanced Virtual Reality Tool for the Assessment of Attention: Normative Study in Spain. *Journal of Attention Disorders*, *20*(6), 542–568. https://doi.org/10.1177/1087054712465335

72. Isik, E. I., Yilmaz, S., Uysal, I., and Basar, S. (2020). Adaptation of the lawton instrumental activities of daily living scale to turkish: Validity and reliability study. *Annals of Geriatric Medicine and Research*, *24*(1), 35–40. https://doi.org/10.4235/agmr.19.0051

73. JASP Team. (2022). JASP (0.16.4.0) [Software]. Retrieved from https://jasp-stats.org/

74. Jeffreys, H. (1961). Theory of Probability, 3rd ed. Oxford University Press.

75. Jerdan, S. W., Grindle, M., van Woerden, H. C., and Kamel Boulos, M. N. (2018). Head-Mounted Virtual Reality and Mental Health: Critical Review of Current Research. *JMIR serious games*, 6(3), e14. https://doi.org/10.2196/games.9226

76. Kaplan, E., Goodglass, H., Weintraub, S., and Goodglass, H. (1983). *Boston naming test*. Lea and Febiger.

77. Kardong-Edgren, S. S., Farra, S. L., Alinier, G., and Young, H. M. (2019). A Call to Unify Definitions of Virtual Reality. *Clinical Simulation in Nursing*. DOI: 10.1016/j.ecns.2019.02.006

78. Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., and Jaffe, M. W. (1963). Studies of Illness in the Aged: The Index of ADL: A Standardized Measure of Biological and Psychosocial Function. *JAMA: The Journal of the American Medical Association*, *185*(12), 914–919. https://doi.org/10.1001/jama.1963.03060120024016

79. Kaufmann, B. C., Cazzoli, D., Pflugshaupt, T., Bohlhalter, S., Vanbellingen, T., Müri, R. M., Nef, T., and Nyffeler, T. (2020). Eyetracking during free visual exploration detects neglect more reliably than paper-pencil tests. *Cortex*, 129, 223–235. https://doi.org/10.1016/j.cortex.2020.04.021

80. Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *Int. J. Aviat. Psychol.* 3, 203–220. doi:10.1207/s15327108ijap0303_3

81. Keysers, C., Gazzola, V., and Wagenmakers, E. J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature neuroscience*, 23(7), 788–799. https://doi.org/10.1038/s41593-020-0660-4

82. Kim, D. Y., Ku, J., Chang, W. H., Park, T. H., Lim, J. Y., Han, K., Kim, I. Y., and Kim, S. I. (2010). Assessment of post-stroke extrapersonal neglect using a three-dimensional immersive virtual street crossing program. *Acta Neurologica Scandinavica*, *121*(3), 171–177. https://doi.org/10.1111/j.1600-0404.2009.01194.x

83. Kisker, J., Gruber, T., and Schöne, B. (2021). Virtual reality experiences promote autobiographical retrieval mechanisms: Electrophysiological correlates of laboratory and virtual experiences. *Psychological research*, *85*(7), 2485–2501. https://doi.org/10.1007/s00426-020-01417-x

84. Knobel, S. E. J., Kaufmann, B. C., Gerber, S. M., Cazzoli, D., Müri, R. M., Nyffeler, T., and Nef, T. (2020). Immersive 3D Virtual Reality Cancellation Task for Visual Neglect Assessment: A Pilot Study. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00180

85. Kowalewski, K. (2016). *Microsoft HoloLens, Darstellung des Benutzersichtfelds bei Raumprojektion* [Photograph]. Wikimedia Commons. https://upload.wikimedia.org/wikipedia/commons/thumb/2/27/MicrosoftHoloLensRoomProjection.JPG/1200px-MicrosoftHoloLensRoomProjection.JPG?20160410190434

86. Kruschke, J. K., and Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic bulletin and review*, *25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

87. Kuula Team (2017). *Tutorial: How to stitch and upload 360 photos from a Samsung Gear 360* [Infographic]. Kuula Blog. https://blog.kuula.co/assets/gear360/header.jpg

88. Lanier, J. (2017). Dawn of the new everything: encounters with reality and virtual reality. First edition. New York, Henry Holt and Company.

89. La Paglia, F., La Cascia, C., Rizzo,R., Riva, G., La Barbera, D. (2012). Assessment of executive functions in patients with obsessive compulsive disorder by neuro VR. *Annual Review of CyberTherapy and Telemedicine*, *10*, 98–102. https://doi.org/10.3233/978-1-61499-121-2-98

90. Lezak, M. D., Howieson, D. B., Bigler, E. D., and Tranel, D. (2012). *Neuropsychological assessment, 5th ed.* Oxford University Press.

91. Lin, J.J., Duh, H.B., Abi-Rached, H., Parker, D.E., and Furness, T.A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. *Proceedings IEEE Virtual Reality* 2002, 164-171.

92. Livingstone, S. A., and Skelton, R. W. (2007). Virtual environment navigation tasks and the assessment of cognitive deficits in individuals with brain injury. *Behavioural Brain Research*, *185*(1), 21–31. https://doi.org/http://dx.doi.org/10.1016/j.bbr.2007.07.015

93. Machado, M. L., Lefèvre, N., Philoxene, B., Le Gall, A., Madeleine, S., Fleury, P., Smith, P. F., and Besnard, S. (2019). New software dedicated to virtual mazes for human cognitive investigations. *Journal of Neuroscience Methods*, *327*, 108388. https://doi.org/10.1016/j.jneumeth.2019.108388

94. Macmillan, N. A. (2002). Signal detection theory. In H. Pashler and J. Wixted (Eds.), Stevens' handbook of experimental psychology: Methodology in experimental psychology (pp. 43–90). John Wiley and Sons Inc.

95. Mantovani, F., Castelnuovo, G., Gaggioli, A., and Riva, G. (2003). Virtual reality training for health-care professionals. *CyberPsychology and Behavior*, 6(4), 389–395. https://doi.org/10.1089/109493103322278772

96. Martelli, D., Prado, A., Xia, B., Verghese, J., and Agrawal, S. K. (2019). Development of a Virtual Floor Maze Test - Effects of Distal Visual Cues and Correlations with Executive Function in Healthy Adults. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, *27*(10), 2229–2236. https://doi.org/10.1109/TNSRE.2019.2938103

97. Matheis, R. J., Schultheis, M. T., Tiersky, L. A., DeLuca, J., Millis, S. R., and Rizzo, A. (2007). Is learning and memory different in a virtual environment? *Clinical Neuropsychologist*, *21*(1), 146–161. https://doi.org/10.1080/13854040601100668

98. Milgram, P., and Kishino, F. (1994). A taxonomy of mixed reality visual displays. IEICE Trans. Inform. Syst. 77, 1321–1329

99. Miller, M.R., Herrera, F., Jun, H. (2020) Personal identifiability of user tracking data during observation of 360-degree VR video. *Sci Rep* **10**, 17404. https://doi.org/10.1038/s41598-020-74486-y

100. Morris, R. G. M., Garrud, P., Rawlins, J. N. P., and O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature*, *297*(5868), 681–683. https://doi.org/10.1038/297681a0

101. Muttaqien W. (2017). *Tutorial: Modeling, UV Unwrapping and Texturing a Mushroom* [Infographic]. Blendernation Blog. https://www.blendernation.com/wp-content/uploads/2017/04/BlenderNation_texturing_mushroom.png

102. Nason, E. E., Trahan, M., Smith, S., Metsis, V., and Selber, K. (2020). Virtual treatment for veteran social anxiety disorder: A comparison of 360° video and 3D virtual reality. *Journal of Technology in Human Services*, 38(3), 288–308. https://doi.org/10.1080/15228835.2019.1692760

103. Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., and Chertkow, H. (2005). The Montreal Cognitive Assessment,

MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

104. Navarro, M. D., Alcañiz, M., Ferri, J., Lozano, J. A., Herrero, N., and Chirivella, J. (2009). Preliminary validation of Ecotrain-Cognitive: A virtual environment task for safe street crossing in acquired brain injury patients with and without unilateral spatial neglect. *Journal of Cybertherapy and Rehabilitation*, *2*(3), 199–203.

105. Negro Cousa, E., Brivio, E., Serino, S., Heboyan, V., Riva, G., and De Leo, G. (2019). New Frontiers for Cognitive Assessment: An Exploratory Study of the Potentiality of 360° Technologies for Memory Evaluation. *Cyberpsychology, Behavior, and Social Networking*, *22*(1), 76–81. https://doi.org/10.1089/cyber.2017.0720

106. Oker A., Pecune F. and Vallverdu J. (2022), Editorial: Virtual reality for neuropsychology and affective cognitive sciences: Theoretical and methodological avenues for studying human cognition. *Front. Virtual Real*. 3:1100387. doi: 10.3389/frvir.2022.1100387

107. Oliveira, C. R., Lopes Filho, B. J. P., Sugarman, M. A., Esteves, C. S., Lima, M. M. B. M. P., Moret-Tatay, C., Irigaray, T. Q., and Argimon, I. I. L. (2016). Development and Feasibility of a Virtual Reality Task for the Cognitive Assessment of Older Adults: The ECO-VR. *The Spanish Journal of Psychology*, *19,* E95. https://doi.org/10.1017/sjp.2016.96

108. Olton, D. S., and Samuelson, R. J. (1976). Remembrance of places passed: Spatial memory in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*(2), 97–116. https://doi.org/10.1037/0097-7403.2.2.97

109. Ouellet, E., Boller, B., Corriveau-Lecavalier, N., Cloutier, S., and Belleville, S. (2018). The Virtual Shop: A new immersive virtual reality environment and scenario for the assessment of everyday memory. *Journal Of Neuroscience Methods*, *303,* 126–135. https://doi.org/10.1016/j.jneumeth.2018.03.010

110. Packard, M. G., and McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of Learning and Memory*, 65(1), 65–72. https://doi.org/10.1006/nlme.1996.0007

111. Page M.J., McKenzie J.E., Bossuyt P.M., Boutron I., Hoffmann T.C., Mulrow C.D., et al. (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *British Medical Journal*, 372(71). https://doi.org/10.1136/bmj.n71

112. Pallavicini, F., Pedroli, E., Serino, S., Dell'Isola, A., Cipresso, P., Cisari, C., and Riva, G. (2015). Assessing Unilateral Spatial Neglect using advanced technologies: The potentiality of mobile virtual reality. *Technology and Health Care*, *23*(6), 795–807. https://doi.org/10.3233/THC-151039

113. Parisi, T. (2015). Learning virtual reality: developing immersive experiences and applications for desktop, web, and mobile. O'Reilly Media, Incorporated, 2015.

114. Parsons, T. D. (2016). *Clinical Neuropsychology and Technology: What's New and How We Can Use It*. Springer International Publishing. https://doi.org/10.1007/978-3-319-31075-6

115. Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*(DEC). https://doi.org/10.3389/fnhum.2015.00660

116. Parsons, T. D., and Barnett, M. (2017). Validity of a Newly Developed Measure of Memory: Feasibility Study of the Virtual Environment Grocery Store. *Journal of Alzheimer's Disease*, *59*(4), 1227–1235. https://doi.org/10.3233/JAD-170295

117. Parsons, T. D., Bowerly, T., Buckwalter, J. G., and Rizzo, A. A. (2007). A Controlled Clinical Comparison of Attention Performance in Children with ADHD in a Virtual Reality Classroom Compared to Standard Neuropsychological Methods. *Child Neuropsychology*, *13*(4), 363–381. https://doi.org/10.1080/13825580600943473

118. Parsons, T. D., and Rizzo, A. A. (2008). Neuropsychological Assessment of Attentional Processing using Virtual Reality. In *Annual Review of CyberTherapy and Telemedicine* 6(1).

119. Pedroli, E., Serino, S., Cipresso, P., Pallavicini, F., Giglioli, I. A. C., Guastafierro, E., Gaggioli, A., Riva, G., Neglect App. Usability of a new application for assessment and rehabilitation of neglect, Contributed paper, in REHAB '15 Proceedings of the 3rd 2015 Workshop on ICTs for improving Patients Rehabilitation Research Techniques, (Lisbona, 01-02 October 2015), ACM, NEW YORK, NY 2015: 139-143. 10.1145/2838944.2838978

120. Pieri, L., Serino, S., Cipresso, P. et al. The ObReco-360°: a new ecological tool to memory assessment using 360° immersive technology. *Virtual Reality* 26, 639–648 (2022). https://doi.org/10.1007/s10055-021-00526-1

121. Pomplun, M., Sichelschmidt, L., Wagner, K., Clermont, T., Rickheit, G., and Ritter, H. (2001). Comparative visual search: A difference that makes a difference. *Cognitive Science, 25*(1), 3–36. https://doi.org/10.1016/S0364-0213(00)00037-9

122. Prithul A, Adhanom IB and Folmer E (2021) Teleportation in Virtual Reality; A Mini-Review. *Front. Virtual Real*. 2:730792. doi: 10.3389/frvir.2021.730792

123. Raspelli, S., Pallavicini, F., Carelli, L., Morganti, F., Poletti, B., Corra, B., Silani, V., and Riva, G. (2011). Validation of a Neuro Virtual Reality-based version of the multiple errands test for the assessment of executive functions. *Annual Review of CyberTherapy and Telemedicine*, *9*(1), 72–76.

124. Rauschnabel, Philipp and Felix, Reto and Hinsch, Christian and Shahab, Hamza and Alt, Florian. (2022). What is XR? Towards a Framework for Augmented and Virtual Reality. *Computers in Human Behavior*. 133. 107289. 10.1016/j.chb.2022.107289.

125. Realdon, O., Serino, S., Savazzi, F., Rossetto, F., Cipresso, P., Parsons, T. D., Cappellini, G., Mantovani, F., Mendozzi, L., Nemni, R., Riva, G., and Baglio, F. (2019). An ecological measure to screen executive functioning in MS: the Picture Interpretation Test (PIT) 360°. Scientific reports, 9(1), 5690. https://doi.org/10.1038/s41598-019-42201-1

126. Reisberg, B., Shulman, M. B., Torossian, C., Leng, L., and Zhu, W. (2010). Outcome over seven years of healthy adults with and without subjective cognitive impairment. Alzheimer's and dementia: *the journal of the Alzheimer's Association*, 6(1), 11–24. https://doi.org/10.1016/j.jalz.2009.10.002

127. Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. (Les problems.). [The psychological examination in cases of traumatic encepholopathy. Problems.]. *Archives de Psychologie*, *28*, 215–285.

128. Riccio, C. A., Reynolds, C. R., Lowe, P., and Moore, J. J. (2002). The continuous performance test: A window on the neural substrates for attention? *Archives of Clinical Neuropsychology*, 17(3), 235–272. https://doi.org/10.1016/S0887-6177(01)00111-1

129. Richard, P., Massenot, L., Besnard, J., Richard, E., Le Gall, D., and Allain, P. (2010). A virtual kitchen to assess the activities of daily life in Alzheimer's disease. *GRAPP 2010 - Proceedings of the International Conference on Computer Graphics Theory and Applications*, 378–383.

130. Ritchie, K., Carrière, I., Howett, D., Su, L., Hornberger, M., O'Brien, J. T., Ritchie, C. W., and Chan, D. (2018). Allocentric and Egocentric Spatial Processing in Middle-Aged Adults at High Risk of Late-Onset Alzheimer's Disease: The PREVENT Dementia Study. *Journal of Alzheimer's disease: JAD*, *65*(3), 885–896. https://doi.org/10.3233/JAD-180432

131. Riva G. (2002). Virtual reality for health care: the status of research. Cyberpsychology and behavior: the impact of the Internet, multimedia and virtual reality on behavior and society, 5(3), 219–225. https://doi.org/10.1089/109493102760147213

132. Riva, G., Wiederhold, B. K., Chirico, A., Di Lernia, D., Mantovani, F., and Gaggioli, A. (2018). Brain and virtual reality: What do they have in common and how to exploit their potential. *Annual Review of CyberTherapy and Telemedicine*, 16, 3–7.

133. Riva G. (2022). Virtual Reality in Clinical Psychology. *Comprehensive Clinical Psychology*, 91–105. https://doi.org/10.1016/B978-0-12-818697-8.00006-6

134. Rizzo, A. S., and Koenig, S. T. (2017). Is clinical virtual reality ready for primetime? *Neuropsychology*, *31*(8), 877–899. https://doi.org/10.1037/neu0000405

135. Rosci, C., Sacco, D., Laiacona, M., and Capitani, E. (2005). Interpretation of a complex picture and its sensitivity to frontal damage: A reappraisal. *Neurological Sciences*, *25*(6), 322–330. https://doi.org/10.1007/s10072-004-0365-6

136. Rouder, J.N., Speckman, P.L., Sun, D. et al. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review* 16, 225–237 (2009). https://doi.org/10.3758/PBR.16.2.225

137. Rouder, J.N., Engelhardt, C.R., McCabe, S. et al. Model comparison in ANOVA. *Psychon Bull Rev* 23, 1779–1786 (2016). https://doi.org/10.3758/s13423-016-1026-5

138. Ruse, S. A., Davis, V. G., Atkins, A. S., Krishnan, K. R. R., Fox, K. H., Harvey, P. D., and Keefe, R. S. E. (2014). Development of a virtual reality assessment of everyday living skills. *Journal of Visualized Experiments*, *86*. https://doi.org/10.3791/51405

139. Sanchez-Vives, Maria V. and Slater, Mel (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience* 6 (4):332-339.

140. Schneider, C. B., Linse, K., Schönfeld, R., Brown, S., Koch, R., Reichmann, H., Leplow, B., and Storch, A. (2017). Spatial learning deficits in Parkinson's disease with and without mild cognitive impairment. *Parkinsonism and Related Disorders*, *36*, 83–88. https://doi.org/10.1016/j.parkreldis.2016.12.020

141. Schöne, B., Wessels, M. and Gruber, T. Experiences in Virtual Reality: a Window to Autobiographical Memory. *Curr Psychol* **38**, 715–719 (2019). https://doi.org/10.1007/s12144-017-9648-y

142. Schultheis, M. T., Himelstein, J., and Rizzo, A. A. (2002). Virtual reality and neuropsychology: Upgrading the current tools. *Journal Of Head Trauma Rehabilitation*, *17*(5), 378–394. https://doi.org/10.1097/00001199-200210000-00002

143. Seo, K., Kim, J., Oh, D. H., Ryu, H., and Choi, H. (2017). Virtual daily living test to screen for mild cognitive impairment using kinematic movement analysis. *PLoS ONE*, *12*(7). https://doi.org/10.1371/journal.pone.0181883

144. Seo, S., Kim, E., Mundy, P., Heo, J., and Kim, K. (Kenny). (2019). Joint Attention Virtual Classroom: A Preliminary Study. *Psychiatry Investigation*, *16*(4), 292–299. https://doi.org/10.30773/pi.2019.02.08

145. Serino, S., Baglio, F., Rossetto, F., Realdon, O., Cipresso, P., Parsons, T. D., Cappellini, G., Mantovani, F., De Leo, G., Nemni, R., and Riva, G. (2017). Picture Interpretation Test (PIT) 360°: An Innovative Measure of Executive Functions. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-16121-x

146. Serino, S., and Repetto, C. (2018). New Trends in Episodic Memory Assessment: Immersive 360° Ecological Videos. *Frontiers in Psychology*, *9*(OCT), 1878. https://doi.org/10.3389/fpsyg.2018.01878

147. Serino, A., Konik, S., Bassolino, M., Serino, S. and Perez-Marcos, D. (2022). Immersive virtual reality for assessment and rehabilitation of deficits of body representations and cognitive functions. *Revue de neuropsychologie*, 14, 15-26. https://doi.org/10.1684/nrp.2022.0700

148. Serrano F. (2016). *Developing an A-Frame Teleport Component* [Infographic]. A-Frame Blog. https://aframe.io/images/blog/teleport-component/background.png

149. Shallice, T., and Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*(2), 727–741. https://doi.org/10.1093/brain/114.2.727

150. Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

151. Spieker, E. A., Astur, R. S., West, J. T., Griego, J. A., and Rowland, L. M. (2012). Spatial memory deficits in a virtual reality eight-arm radial maze in schizophrenia. *Schizophrenia Research*, *135*(1–3), 84–89. https://doi.org/10.1016/j.schres.2011.11.014

152. Squire, L. R. (2009). Memory and brain systems: 1969-2009. *Journal of Neuroscience*, 29(41), 12711–12716. https://doi.org/10.1523/JNEUROSCI.3575-09.2009

153. Stark, L., Choi, Y., Yu, Y. (1997). Visual Imagery and Virtual Reality. In: Lakshminarayanan, V. (eds) Basic and Clinical Applications of Vision Science. Documenta Ophthalmologica Proceedings Series, vol 60. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-5698-1_29

154. Stein, J.-P., and Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition, 160,* 43–50. https://doi.org/10.1016/j.cognition.2016.12.010

155. Stewart, Robert. Subjective cognitive impairment. *Current Opinion in Psychiatry* 25(6):p 445-450, November 2012. | DOI: 10.1097/YCO.0b013e3283586fd8

156. Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651

157. Stuchlik, A., Kubik, S., Vlcek, K., and Vales, K. (2014). Spatial navigation: Implications for animal models, drug development and human studies. *Physiological Research*, 63(SUPPL). https://doi.org/10.33549/physiolres.932660

158. Swanson J., Schuck S., Mann M., Carlson C., Hartman K., Sergeant J.A. (2001) Over- identification of extreme behavior in evaluation and diagnosis of ADHD/HKD. www.adhd.net/SWAN_Paper.pdf.

159. Tarnanas, I., Schlee, W., Tsolaki, M., Müri, R., Mosimann, U., and Nef, T. (2013). Ecological Validity of Virtual Reality Daily Living Activities Screening for Early Dementia: Longitudinal Study. *JMIR Serious Games*, *1*(1), e1. https://doi.org/10.2196/games.2778

160. Tascòn, Laura, Garcia-Moreno, M.., Cimadevilla, L. and Manuel, J. (2017). Almeria spatial memory recognition test (ASMRT): Gender differences emerged in a new passive spatial task. *Neuroscience Letters*, *651*(NA), 188–191. https://doi.org/10.1016/j.neulet.2017.05.011

161. Tascòn, L., Castillo, J., Cimadevilla, L., and Manuel, J. (2019). Age-related differences in the elderly in a spatial recognition task. *Memory*, *27*(10), 1415–1422. https://doi.org/10.1080/09658211.2019.1663216

162. Tatler, B. W., and Melcher, D. (2007). Pictures in mind: initial encoding of object properties varies with the realism of the scene stimulus. *Perception*, 36(12), 1715–1729. https://doi.org/10.1068/p5592

163. Tosi, G., Borsani, C., Castiglioni, S., Daini, R., Franceschi, M., and Romano, D. (2020). Complexity in neuropsychological assessments of cognitive impairment: A network analysis approach. *Cortex, 124*, 85–96. https://doi.org/10.1016/j.cortex.2019.11.004

164. Trés, E. S., and Brucki, S. (2014). Visuospatial processing: A review from basic to current concepts. *Dementia and neuropsychologia*, *8*(2), 175–181. https://doi.org/10.1590/S1980-57642014DN82000014

165. Tupper, D. E., and Cicerone, K. D. (2012). *The neuropsychology of everyday life: Assessment and basic competencies (Vol. 2)*. Springer Science and Business Media.

166. Tuena, C., Pedroli, E., Trimarchi, P. D., Gallucci, A., Chiappini, M., Goulene, K., Gaggioli, A., Riva, G., Lattanzio, F., Giunco, F., and Stramba-Badiale, M. (2020). Usability Issues of Clinical and Research Applications of Virtual Reality in Older People: A Systematic Review. *Frontiers in human neuroscience*, *14*, 93. https://doi.org/10.3389/fnhum.2020.00093

167. Vallejo, V., Wyss, P., Chesham, A., Mitache, A. V, Mueri, R. M., Mosimann, U. P., and Nef, T. (2017). Evaluation of a new serious game-based multitasking assessment tool for cognition and activities of daily living: Comparison with a real cooking task. *Computers In Human Behavior*, *70*, 500–506. https://doi.org/10.1016/j.chb.2017.01.021

168. Vandenberg, S. G., and Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604. https://doi.org/10.2466/pms.1978.47.2.599

169. VandenBos, G. R. (Ed.). (2015). *APA dictionary of psychology*, *2nd ed.* American Psychological Association. https://doi.org/10.1037/14646-000

170. van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., and van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, 85(3), 842–860. https://doi.org/10.1111/cdev.12169

171. Vasconcelos-Raposo, José and Melo, Miguel and Teixeira, Carla and Cabral, Luciana and Bessa, Maximino. (2018). Adaptation and Validation of the ITC - Sense of Presence Inventory for the Portuguese language. *International Journal of Human-Computer Studies*. 125. 10.1016/j.ijhcs.2018.12.005.

172. Ventura, S., Baños, R. M., and Botella, C. (2018). Virtual and Augmented Reality: New Frontiers for Clinical Psychology. In Nawaz, M. (Ed.) *State of the Art Virtual Reality and Augmented Reality Knowhow*. InTech. https://doi.org/10.5772/intechopen.74344

173. Wall, K. J., Cumming, T. B., Koenig, S. T., Pelecanos, A. M., and Copland, D. A. (2018). Using technology to overcome the language barrier: the Cognitive Assessment for Aphasia App. *Disability and Rehabilitation*, *40*(11), 1333–1344. https://doi.org/10.1080/09638288.2017.1294210

174. Wang, D., Guo, Y., Liu, S., Zhang, Y., Xu, W., and Xiao, J. (2019). Haptic display for virtual reality: Progress and challenges. *Virtual Real. Intelligent Hardw*. 1 (2), 136–162. doi:10.3724/sp.j.2096-5796.2019.0008

175. Weniger, G., Ruhleder, M., Lange, C., Wolf, S., and Irle, E. (2011). Egocentric and allocentric memory as assessed by virtual reality in individuals with amnestic mild cognitive impairment. *Neuropsychologia*, *49*(3), 518–527. https://doi.org/10.1016/j.neuropsychologia.2010.12.031

176. Westgate M. J. (2019) revtools: An R package to support article screening for evidence synthesis. *Research Synthesis Methods* http://doi.org/10.1002/jrsm.1374

177. Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., and Wagenmakers, E. J. (2011). Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. *Perspectives on psychological science*, 6(3), 291–298. https://doi.org/10.1177/1745691611406923

178. Wilkins, L. K., Girard, T. A., King, J., King, M. J., Herdman, K. A., Christensen, B. K., and King, J. (2013). Spatial-memory deficit in schizophrenia spectrum disorders under viewpoint-independent demands in the virtual courtyard task. *Journal of Clinical and Experimental Neuropsychology*, *35*(10), 1082–1093. https://doi.org/http://dx.doi.org/10.1080/13803395.2013.857389

179. Wilson, B., Cockburn, J. and Halligan, P. (1987). Development of a behavioral test of visuospatial neglect. *Archives of Physical Medicine and Rehabitation*, 68(2), 98–102.

180. Witmer, B. G., and Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence Teleoperat. Virt. Environ*. 7, 225–240. doi: 10.1162/105474698565686

181. Yan, B., Ni, S., Wang, X., Liu, J., Zhang, Q., and Peng, K. (2020). Using Virtual Reality to Validate the Chinese Version of the Independent Television Commission-Sense of Presence Inventory. SAGE Open, 10(2). https://doi.org/10.1177/2158244020922878

182. Yuan, Y., and Johnson, V. E. (2008). Bayesian hypothesis tests using nonparametric statistics. *Statistica Sinica*, 18(3), 1185–1200. http://www.jstor.org/stable/24308537

183. Zhang, L., Abreu, B. C., Seale, G. S., Masel, B., Christiansen, C. H., and Ottenbacher, K. J. (2003). A virtual reality environment for evaluation of a daily living skill in brain injury rehabilitation: reliability and validity. *Archives of Physical Medicine and Rehabilitation*, *84*(8), 1118–1124. https://doi.org/10.1016/s0003-9993(03)00203-x

184. Zimmermann, P. and Fimm, B. (2004). A test battery for attentional performance. In *Applied Neuropsychology of Attention* (pp. 124–165). Psychology Press. https://doi.org/10.4324/9780203307014-12

185. Zhou, NN., Deng, YL. Virtual reality: A state-of-the-art survey. *Int. J. Autom. Comput.* 6, 319–325 (2009). https://doi.org/10.1007/s11633-009-0319-9

186. Zulueta, A., Diaz-Orueta, U., Crespo-Eguilaz, N., and Torrano, F. (2019). Virtual Reality-based Assessment and Rating Scales in ADHD Diagnosis. *Psicologia Educativa*, *25*(1), 13–22. https://doi.org/10.5093/psed2018a18

Appendix

Screenshots showing the scenarios included in the MS-360°



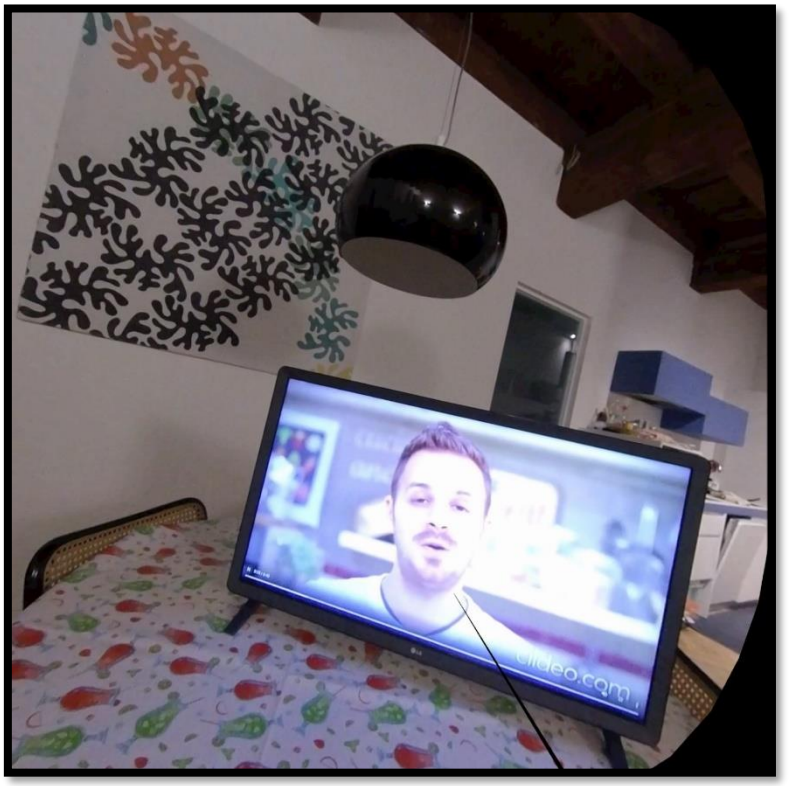
1. Free Visual Exploration Task



2. Target Selection Task



3. Object Naming



4. Object Recall (Immediate)



5. Story Recall (Immediate)



6. Word Reading

7. Semantic Fluency

9. Action Planning



8. Cognitive Estimates



10. Written Comprehension



11. Sustained Attention



12. Object Recognition