UNIVERSITÀ CATTOLICA DEL SACRO CUORE

Sede di Milano

Dottorato di ricerca in Psicologia

Ciclo XXXV

S.S.D. M-PSI/03

EXecutive-functions Innovative Tool - EXIT 360: Development and Validation of a new 360-video instrument for executive functions

Coordinatore:

Ch.mo Prof. Camillo Regalia

Tesi di Dottorato di:

Francesca Borgnis

N. Matricola: 4912505

Anno Accademico 2021/2022

UNIVERSITÀ CATTOLICA DEL SACRO CUORE

Sede di Milano

Dottorato di ricerca in Psicologia

Ciclo XXXV

S.S.D. M-PSI/03

EXecutive-functions Innovative Tool - EXIT 360: Development and Validation of a new 360-video instrument for executive functions

Coordinatore:

Ch.mo Prof. Camillo Regalia

Tutor:

Ch.mo Prof. Pietro Cipresso

Tesi di Dottorato di:

Francesca Borgnis

N. Matricola: 4912505

Anno Accademico 2021/2022

Abstract

This thesis offers a broad overview of EXecutive-functions Innovative Tool 360° (EXIT 360°) development from its concept to validation as a sensitive new-brand instrument for evaluating executive functionality. As described in **Chapter 1** ("*Executive Functions and Virtual Reality*"), among cognitive impairments, executive dysfunction represents a significant and increasing public health problem due to its high negative impact on daily activities (e.g., shopping, cooking, managing money, working) and quality of life. Therefore, identifying early strategies to detect these executive impairments appears to be a priority in the field of neuropsychology. The overt ecological limitations of traditional paper-and-pencil neuropsychological tests and multiple challenges associated with administering tests in real-life scenarios have led to the growing use of new technological solutions, especially Virtual Reality (VR) tools, for assessing executive functions in real-life contexts. Recently, advances in 360° technology emerged as a valuable alternative approach to create VR–immersive applications at a low cost. In this framework, we conceptualized, designed, and developed EXecutive-functions Innovative Tool 360° (EXIT 360°), a new 360°-based instrument for an ecologically valid and multicomponent assessment of executive functioning (Borgnis, Baglio, Pedroli, Rossetto, Riva, et al., 2021). **Chapter 2** ("*EXIT 360° - From Concept to Validation*") describes the concept, design, and development of EXIT 360°, focusing on its main innovative characteristics. EXIT 360° consists of a new task for executive functions delivered via a smartphone and a comfortable head-mounted display. EXIT 360° engages participants in a "game for health" where they must perform seven everyday subtasks of increasing complexity in 360° domestic environments. **Chapter 3** ("*EXIT 360° - Usability Studies*") shows the results of the usability studies involving patients with PD and healthy control volunteers. Parkinson's Disease (PD) was chosen among numerous clinical conditions showing executive impairments since it is well known that executive dysfunction represents a common non-motor symptom in early-stage non-demented PD with obvious adverse consequences for daily functioning and quality of life. Briefly, the results show that healthy controls (including elderly subjects) and patients with PD consider EXIT 360° as a straightforward, pleasant, engaging, usable, and easy-to-learn technological instrument and claim to have had an excellent experience using it (Borgnis, Baglio, Pedroli, Rossetto, Isernia, et al., 2021; Borgnis, Baglio, Pedroli, Rossetto, Meloni, et al., 2022). **Chapter 4** ("*EXIT 360° - Convergent Validity Studies*") assesses the convergent validity of the tool. Data shows an excellent convergent validity of EXIT 360°, displaying a strong positive correlation with standardized traditional neuropsychological

instruments for executive functioning. The multicomponent dimension appears critical in evaluating executive functioning since it is a complex and heterogeneous construct involving a wide range of cognitive processes and behavioural skills responsible for many everyday activities (Chan et al., 2008). Finally, the effectiveness of EXIT 360° in discriminating between pathological and control groups will be examined in **Chapter 5** ("*EXIT 360° - Construct Validity and Diagnostic Assessment*"). Classification analysis confirms the great potential of the EXIT 360° for distinguishing between PwPD and controls with high precision (≥ 79%). EXIT 360° scores also show higher diagnostic accuracy in predicting PD group membership compared to traditional neuropsychological tests. In this context, EXIT 360° provides a new paradigm in which patients are active participants within an ecological virtual world (Parsons, 2015; Riva, 2009), where it is possible to simulate life-like challenges that reproduce everyday situations and, as a result, actual patient's executive status.

In conclusion, the studies reported in this thesis have tried to answer three questions: [1] *"Will the EXIT 360 ° technology be usable by both the people with low technological expertise, elderly and patients with cognitive dysfunction?"*; [2] *"Will EXIT 360° actually evaluate multiple components of executive functioning?"*; [3] *"Will EXIT 360° be able to discriminate between patients and healthy subjects with good diagnostic accuracy?"*. Findings offer clear evidence that EXIT 360° must be seen as a valuable and innovative instrument for an ecologically valid and multicomponent evaluation of executive functioning, highly usable for prompt diagnosis and early patient enrolment in focused rehabilitation. I strongly believe that this innovative 360°-based tool, which is simply usable in clinical settings, has the potential to transform patients' and clinicians' evaluation experience radically.

Table of Contents

Introduction **1**

1. Executive Functions and Virtual Reality **3**

1.1 Executive Functions: What are they? 3

1.2 Executive Function and Brain: What is the link? 4

1.3 Executive Functions in everyday situations: What is the impact? 5

1.4 Executive Functions: How are they traditionally evaluated? 6

1.5 Executive Functions: Virtual Reality as new frontiers for assessment 7

1.5.1 Executive Dysfunction in Psychiatric and Neurologic population: What did virtual reality discover? 10

1.6 Executive Functions and 360° Technology: A new opportunity? 14

1.7 Why have we chosen Parkinson's Disease as a clinical population? 15

2. EXIT 360° - From Concept to Validation **19**

2.1 CONCEPT, DESIGN AND DEVELOPMENT 19

2.1.1 EXIT 360°: Virtual Environments 19

2.1.2 EXIT 360°: Seven Subtasks 21

2.1.3 EXIT 360°: Evaluation Procedure 25

2.1.4 EXIT 360°: Outcome measures 26

2.2 VALIDATION PROCEDURE 27

3. EXIT 360° - Usability Studies **31**

Study 1: EXIT 360° Usability and User Experience in Healthy Control Subjects 33

Material and Methods 33

Results 34

Study 2: EXIT 360° Usability and User Experience in Parkinson's Disease 41

Material and Methods 41

Results 42

Conclusion 50

Usability Studies: Conclusion, Limitations and Future Perspectives 51

4. EXIT 360° - Convergent Validity Study **55**

Material and Methods 57

Results 61

Discussion and Conclusion 65

5. EXIT 360° - Construct Validity and Diagnostic Assessment **69**

Material and Methods 70

Results 73

Discussion and Conclusion 81
6. Discussion and Conclusion 85
References 93

Introduction

Neuropsychology is a specified field within psychology dedicated to studying the relationships between the brain and behaviour, focusing on diagnosing brain disorders, assessing cognitive and behavioural functioning, and planning efficient rehabilitative treatments. In this framework, neuropsychological assessment is traditionally viewed as an integral part of the neurological examination and involves normatively informed applying performance-based evaluations of different cognitive domains, such as memory, language, visuospatial abilities, and executive functions. An early and comprehensive neuropsychological assessment allows for determining patients' cognitive strengths and weaknesses. As a result, neuropsychologists will be able to accurately detect cognitive impairments that need to be addressed in rehabilitation, allowing them to tailor the rehabilitative treatments to patients' needs.

Among possible cognitive impairments, executive dysfunction represents a significant and increasing public health problem due to its high negative impact on daily activities (e.g., shopping, cooking, managing money, working) and quality of life. Therefore, identifying early strategies to detect these executive impairments appears to be a priority, also to plan timely rehabilitation programs aimed at enhancing daily functioning.

The overt ecological limitations of traditional paper-and-pencil neuropsychological tests and multiple challenges associated with administering tests in real-life scenarios have led to the growing use of new technological solutions, especially Virtual Reality (VR) tools, for assessing executive functions in real-life contexts. Recently, advances in 360° technology emerged as a valuable alternative approach to create VR–immersive applications at a low cost. In this framework, we conceptualized, designed, and developed EXecutive-functions Innovative Tool 360° (EXIT 360°), a new 360°-based instrument for an ecologically valid and multicomponent assessment of executive functioning (Borgnis, Baglio, Pedroli, Rossetto, Riva, et al., 2021).

This thesis offers a broad overview of EXIT 360° development from its concept to validation as a sensitive new-brand instrument for evaluating executive functionality. EXIT 360° consists of a new task for executive functions delivered via a smartphone and a comfortable head-mounted display. EXIT 360° engages participants in a "game for health" where they must perform seven everyday subtasks of increasing complexity in 360° domestic environments. Chapter 2 will thoroughly describe the concept, design, and development of EXIT 360°, focusing on its main innovative characteristics.

The following chapters will focus on the main psychometric properties of EXIT 360°, namely usability, convergent validity, and construct validity, to answer three questions: [1] *"Will the EXIT 360 ° technology be usable by both the people with low technological expertise, elderly and patients with cognitive dysfunction?"*; [2] *"Will EXIT 360° actually evaluate multiple components of executive functioning?"*; [3] *"Will EXIT 360° be able to discriminate between patients and healthy subjects with good diagnostic accuracy?"*. Parkinson's Disease (PD) was chosen among numerous clinical conditions showing executive impairments since it is well known that executive dysfunction represents a common non-motor symptom in early-stage non-demented PD with obvious adverse consequences for daily functioning and quality of life.

The major findings of the usability studies involving patients with PD and healthy control volunteers will be covered in Chapter 3. Additionally, Chapter 4 will detail the investigation findings to evaluate the convergent validity. Finally, the effectiveness of EXIT 360° in discriminating between pathological and control groups will be examined in Chapter 5.

Overall, this thesis will offer clear evidence that EXIT 360° can be seen as a valuable and innovative instrument for an ecologically valid and multicomponent evaluation of executive functioning, highly usable for prompt diagnosis and early patient enrolment in focused rehabilitation. I strongly believe that this innovative 360°-based tool could have the potential to transform patients' and clinicians' evaluation experience radically.

1. Executive Functions and Virtual Reality

Neuropsychological assessment has been historically seen as an integral part of the neurological examination and entails normatively informed applying performance-based assessments of different cognitive skills (Harvey, 2012). Among these cognitive abilities, the evaluation of executive functioning represents a challenge for neuropsychologists due to the construct's heterogeneity (Diamond, 2013; Stuss & Alexander, 2000) and the methodological issues (Barker et al., 2004; Chaytor & Schmitter-Edgecombe, 2003; Goldstein, 1996).

1.1 Executive Functions: What are they?

"Executive function" (EF) is a complex and heterogeneous construct involving a wide range of cognitive processes and behavioural skills responsible for controlling and regulating actions (e.g., starting and stopping activities or monitoring) (Burgess & Simons, 2005; Chan et al., 2008) and carrying out complex or non-routine tasks (e.g., ability to perform two tasks simultaneously) (M. K. Alderman, 2013; Alvarez & Emory, 2006; Godefroy, 2003). Over the years, various authors have offered different definitions of EFs without reaching a complete agreement (Baddeley et al., 1997; Chan et al., 2008; Diamond, 2013; Lezak, 1995). Lezak formulated one of the first definitions of EFs in 1995, according to which EFs involve four components: 1) Volition - willingness and capacity to establish a goal and begin the performance; 2) Effective performance - the capacity to monitor, adjust, and manage a specific activity; 3) Planning - the capacity to identify and organize the steps and objects essential to achieve a goal or activity and 4) Intentional execution of the planned activity that requires the start, supervision and modification or suspension of a sequence of actions when necessary (Lezak, 1995). Other authors have subsequently expanded the concept of EFs, referring to them as a set of mental abilities that control the activity of other cognitive systems involved in the execution of goal-directed behaviour, such as memory and reasoning (Baddeley et al., 1997).

Another interesting definition of EF was formulated by Chan and colleagues in 2008. The authors described EFs as "an umbrella term comprising a wide range of cognitive processes and behavioural competencies which include verbal reasoning, problem-solving, planning, sequencing, the ability to sustain attention, resistance to interference, utilisation of feedback, multitasking, cognitive flexibility, and the ability to deal with the novelty" (Chan et al., 2008). Later, other authors decided to reduce this list to three essential cognitive constructs (i.e., inhibitory control, working memory and cognitive flexibility), which serve as the foundation for all higher-order EFs, such as reasoning, problem-solving and planning (Diamond, 2013).

1.2 Executive Function and Brain: What is the link?

Burgess and Simons (2005) and Chan and colleagues (2008) showed a lack of consensus on the functional and cerebral architecture of executive functioning (Burgess & Simons, 2005; Chan et al., 2008). The hypotheses range from the unitary nature of the executive system to multiple and more diverse systems (Duncan, 1986; Luria, 1973; Luria et al., 1966; Norman & Shallice, 1986; Stuss & Benson, 1984). For a long time, the existence of cognitive abilities supported by the frontal lobes was contested (Benton, 1991). Only around the 1960s did evidence emerge about frontal injury's effects on humans and primates.

Figure 1.1: Frontal Lobe

The crucial frontal lobe is said to be involved in supporting executive processes involved in many real-life situations, such as preparing meals, managing money, shopping for groceries, doing housework, or using a cell phone. This belief is supported by the cognitive neuroscience of EFs, which has been rapidly developing in recent years and is being driven by technological progress (for example, new brain imaging methods and better lesion location methods) (Burgess et al., 2006). In the healthy elderly population, problems affecting EFs are frequently linked to the frontal cortex's aging (Burke & Barnes, 2006; Raz, 2000). However, we refer to "Frontal Syndrome" or "Dysexecutive Syndrome" (Bechara et al., 1994; Robertson et al., 1997) after lesions localized in one or more areas of the frontal and/or prefrontal cortex as a result of acute damage like traumatic brain injury and stroke (Baddeley & Wilson, 1988; Nys et al., 2007) or neurodegenerative pathologies such as Parkinson's disease (Aarsland et al., 2005, 2007; Kudlicka et al., 2011) and Multiple Sclerosis (Nebel et al., 2007). Despite the crucial role of the frontal cortex, several studies have shown that an alteration in EFs can be consequential to damage to other areas of the brain since the frontal regions have connections with many other cortical and subcortical areas, such as the amygdala, cerebellum, and basal ganglia (for a review, see (Tekin & Cummings, 2002)).

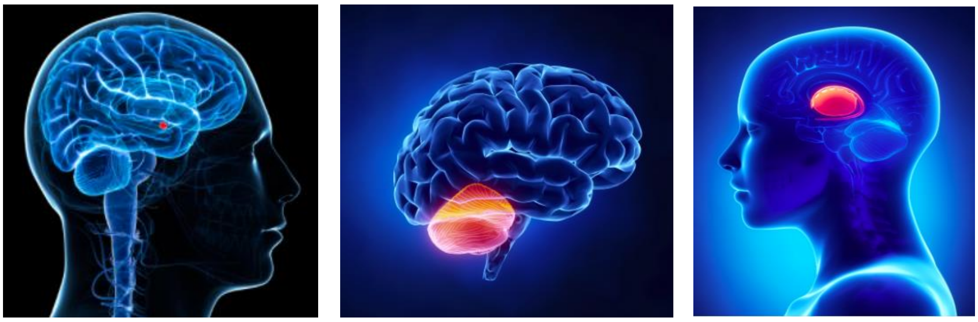


Figure 1.2: Cortical and subcortical areas connected with the frontal regions: Amygdala, Cerebellum and Basal Ganglia, respectively

1.3 Executive Functions in everyday situations: What is the impact?

As previously said, alterations in EFs can negatively impact various mental and behavioural processes, such as difficulty performing two tasks simultaneously, a propensity to get distracted easily, trouble starting or stopping activities, and reduced ability to learn new tasks (M. K. Alderman, 2013). Furthermore, they can impair the ability to develop new strategies and plan, monitor or inhibit irrelevant stimuli and responses (Crawford, 1998). It is clear that deficits in executive functioning could affect the ability to perform numerous activities of daily living, both in the "basic activities" (e.g., personal hygiene, dressing, eating) and in those “instrumental” (e.g., preparing meals, managing money, using the mobile phone, shopping, or housework) (Diamond, 2013; Fortin et al., 2003; Katz, 2011; Vaughan & Giovanello, 2010). Therefore, the Dysexecutive Syndrome, typical of many psychiatric and neurological pathologies, constitutes a significant public health issue due to its high impact on personal independence, ability to work, educational success, social relationships, and cognitive and psychological development (Diamond, 2013; Goel et al., 1997; Green, 1996, 2006), with unavoidable consequences on a person’s quality of life and feelings of personal wellbeing (Gitlin et al., 2001).

Therefore, identifying early strategies for evaluating executive dysfunction appears to be a priority to minimise the effects of executive impairments and improve everyday functioning (Levine et al., 2007).

1.4 Executive Functions: How are they traditionally evaluated?

EFs are traditionally evaluated through laboratory tasks or conventional paper-and-pencil neuropsychological tests, which guarantee standardized procedures and scores that make them valid and reliable. The Modified Wisconsin Card Sorting Test (Nelson, 1976), the Stroop Test (Stroop, 1935), and the Trail Making Test (TMT) (Reitan, 1992a) are perfect examples of neuropsychological tests specific for EFs. Over the years, a growing variety of tests have been created to evaluate various patients (Chan et al., 2008). An assessment protocol may include a unique task for assessing a single cognitive process, such as the Tower of London for problem-solving skills (Allamanno et al., 1987), or tests batteries to evaluate several aspects of executive functionality, such as the Frontal Assessment Battery (FAB) (Appollonio et al., 2005; Dubois et al., 2000).

These traditional paper-pencil tests are highly structured, with clearly standardized procedures and goals but appear not to be able to reliably predict the complexity of executive functioning in real-life settings (Burgess et al., 2006; Chan et al., 2008; Chaytor & Schmitter-Edgecombe, 2003; Klinger et al., 2004; Shallice & Burgess, 1991).

Indeed, multiple studies found that many patients with Dysexecutive Syndrome normally perform on conventional neuropsychological tests while nevertheless experiencing significant difficulties with day-to-day activities (Shallice & Burgess, 1991). This issue might be brought about by the classic EF tests' lack of ecological validity (Chan et al., 2008), which makes it appear that they cannot adequately represent the complexity of EFs in a realistic setting. Standard paper and pencil tests require simple responses to a unique event, while some day-to-day tasks could call for a more complicated series of answers (Chan et al., 2008). In other words, most of the situations experienced outside the clinic differ from those in which patients do the assessment (that is, they show little "representativeness").

Over the years, several works have supported the necessity to use instrumental daily life activities and everyday contexts in the assessment to understand how executive deficits may affect daily functioning (Burgess et al., 2006; Manchester et al., 2004). In this framework, Burgess and colleagues suggested neuropsychological evaluations based on models generated from directly observable daily behaviors (Burgess et al., 2006). This innovative "function-led" approach differentiates from other methods that emphasize abstract cognitive "constructs" by focusing on the role of EFs within the complexity of the "functional" behaviors present in real-life situations. This novel approach may evaluate whether patients can efficiently manage and orient cognitive resources within the complexity of the outside environment, allowing a deeper comprehension of the patient's neuropsychological profile and future personalized rehabilitation (Pedroli et al., 2016).

To overcome the ecological issue, clinicians and researchers paid attention to develop tests able to evaluate executive functioning in real-life settings (Chaytor & Schmitter-Edgecombe, 2003; Jurado & Rosselli, 2007), such as the Multiple Errands Test (MET) (N. Alderman et al., 2003; Shallice & Burgess, 1991) and Behavioral Assessment of the Dysexecutive Syndrome (BADS) (Wilson et al., 1997). For instance, the MET assessed patients while performing various daily tasks (such as shopping) in a real supermarket while adhering to set guidelines and timing requirements. In contrast to laboratory settings, assessing EFs in real-world scenarios permitted a more accurate estimate of the patient's deficits (Rand et al., 2009). However, this method has additional drawbacks, including extended study times, high costs, organizational challenges (such as obtaining permission from local companies), and poor controllability of the experimental condition or applicability with patients who have significant behavioural, psychiatric, and motor deficits (Bailey et al., 2010; Cherniack, 2011).

The ecological limitations of conventional neuropsychological batteries and difficulties in administering tests in real-life scenarios have paved the way for using technological tools such as Virtual Reality (VR) to assess EFs in real-life (Bohil et al., 2011; Neguț, 2014; Neguț et al., 2016).

1.5 Executive Functions: Virtual Reality as new frontiers for assessment

Interactive technologies are quickly developing into a promising tool for simulating a real environment, with extensive use in health care, including neuropsychological assessment. Examples of interactive technologies include virtual reality, mobile devices and sensors, serious games, and 360° video. VR is a type of human-computer interface that enables to create realistic spatial and temporal scenarios, situations or objects that could allow an ecologically valid evaluation of EFs by replicating conditions of daily life (Bohil et al., 2011; Campbell et al., 2009; Parsons, 2015). With the aid of VR-based solutions, clinicians may be able to observe their patients while performing everyday executive tasks (e.g., shopping) in ecologically controlled environments like supermarkets and kitchens, "like in real life" (Figure 1.3) (Climent et al., 2010; Maggio et al., 2018, 2019).

Figure 1.3: Screenshot of Virtual Multiple Errands (Cipresso et al., 2014)

Additionally, virtual environments in neuropsychological examinations may provide better perceptual environment control, a more accurate stimulus presentation, greater applicability, user-friendly interfaces, data collection and analysis in real-time (M. K. Alderman, 2013; Armstrong et al., 2013; Cipresso, Meriggi, et al., 2013; Parsons, 2015; Parsons et al., 2011; Riva, 2009; Rizzo et al., 2001).

A systematic review (2022) conducted by Borgnis and colleagues revealed that VR can be considered an innovative and useful alternative to evaluate EFs in real-world scenarios (Borgnis, Baglio, Pedroli, Rossetto, Uccellatore, et al., 2022).

Research showed that assessment in real-life scenarios provides a more accurate estimate of the patient's impairments, showing difficulties invisible to traditional measurements (Armstrong et al., 2013; Cipresso, la Paglia, et al., 2013; Parsons, 2015; Rand et al., 2009). To date, thirty VR-based assessment tools have been developed for executive dysfunctions, showing significant variability in implemented environments (e.g., supermarket, kitchen, office, library), stimuli, and tasks (e.g., cooking and shopping). The virtual supermarket appeared as the most used VE (e.g., VMET), and consequently, shopping was chosen as the best task since it entails several activities and actions that require the involvement of a large number of EFs (planning, multitasking, problem-solving, and set-shifting). Overall, participants are immersed in everyday settings where they must complete several tasks involving complex real-life situations requiring subjects to use several EFs (Nir-Hadad et al., 2017), reflecting the cognitive demands of daily functioning (Chaytor & Schmitter-Edgecombe, 2003) (Figure 1.4).

Figure 1.4: Screenshot of Virtual Multiple Errands (Raspelli et al., 2012).
© 2012 by the Massachusetts Institute of Technology

The high flexibility and programmability of VR appeared as crucial characteristics in an evaluation tool as they ensure a controlled and accurate presentation of a wide range of stimuli and distractions/stressors that patients may encounter in daily life (Armstrong et al., 2013). Additionally, flexibility is a desirable quality since it enables contexts to be modified by demographic or clinical characteristics, like age, education, or overall cognitive performance (Renison et al., 2012).

Among critical components of the assessment instrument, the systematic review has also focused on the psychometric properties of tools, including construct validity, convergent validity, usability, and test re-test reliability. Overall, most revised VR-based assessment tools (77%) exhibit strong construct validity, with significant correlations between existing standardized paper-and-pencil tests and the primary outcome measures. On the other hand, to date, only half of the revised designed instruments have demonstrated good construct validity. It should be highlighted that the lack of construct validity constitutes a significant restriction on the use of these tools because their introduction into clinical practice is impossible in the absence of data on diagnostic specificity and sensitivity in clinical populations. Finally, the review marked attention to the paucity of research on usability and test-retest reliability, two crucial elements for an instrument exploitable in a clinical environment. Numerous studies have demonstrated how important it is to evaluate usability and user experience when creating VR-based tools (Pedroli et al., 2013, 2019; Sauer et al., 2020; Serino et al., 2020; Tuena et al., 2020). Usability issues or cybersickness (that lead to user unpleasant experiences) could affect performance, significantly decreasing the test results' validity (Armstrong et al., 2013).

A further systematic review (2022) on VR-based assessment of EFs in psychiatric and clinical populations showed that several studies converged in supporting the feasibility and effectiveness of VR-based tools in the ecologically valid evaluation of executive functionality in psychiatric and neurologic populations (Borgnis et al., under review) (Figure 1.5).

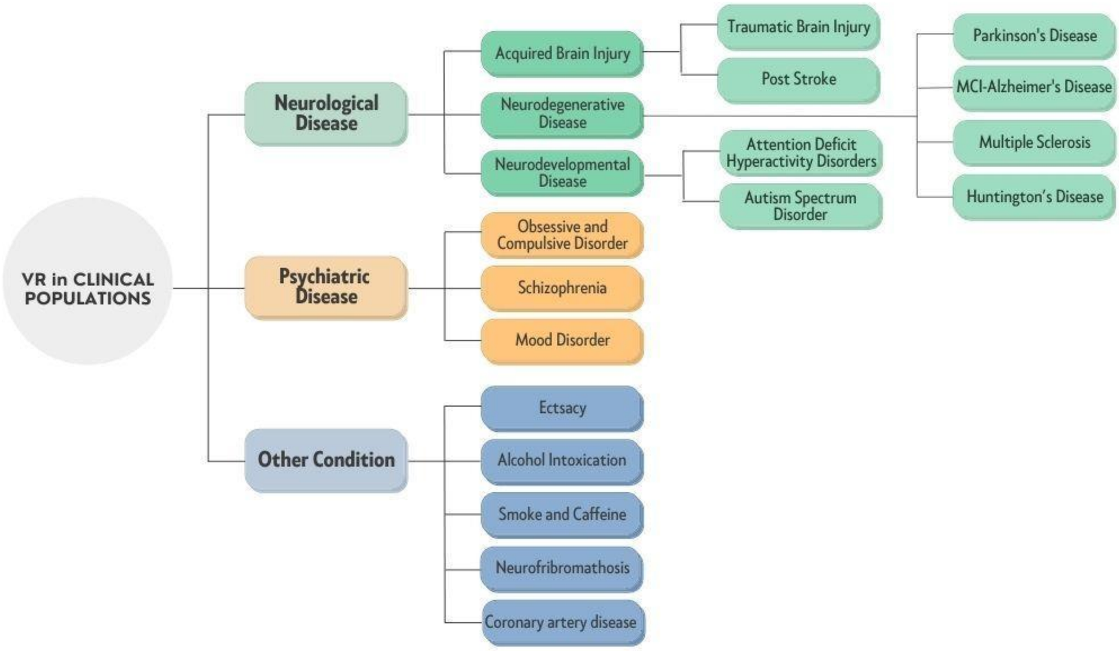


Figure 1.5: Clinical conditions for which VR-based tools have been implemented

These VR-based tools have shown to be not only more environmentally friendly but also more effective than classic paper and pencil tasks: what is remarkable is that this effectiveness is specified for the pathology analyzed and for the EF activated. Additionally, VR-based techniques appeared useful in diagnosing executive dysfunctions in neurodegenerative diseases from the earliest stages.

1.5.1 Executive Dysfunction in Psychiatric and Neurologic population: What did virtual reality discover?

The systematic review showed that schizophrenia, Acquired Brain Injury (ABI - including traumatic brain injury and post-stroke), and Parkinson's Disease appeared as the clinical categories in which there is a significant number of studies on VR-based tools (Borgnis et al., under review). All VR-based tools have been categorized based on three main clinical categories: Psychiatric disorders, Neurological diseases and Other conditions.

Psychiatric disorders: show impairments in the frontal cortex and its connections with subcortical structures with consequent significant EF deficits. These cognitive deficits affect patients' everyday functioning, worsening the clinical course of the disease (Aubin et al., 2009). For example, executive deficits are related to suicidal behavior in bipolar disorder patients or a lack of adherence to treatment and underperformance in the workplace and educational settings in several psychiatric conditions (e.g., bipolar disorder, schizophrenia, obsessive-compulsive disorder, and drug addiction). To date, studies have focused on developing VR-based tools for evaluating executive impairments in schizophrenia, Obsessive and Compulsive Disorders (OCD), and Mood Disorders, showing promising results (Table 1.1).

Psychiatric Disorders	
Schizophrenia	Virtual Action Planning Supermarket (VAP-S) Virtual Supermarket Shopping Task (VSST) Virtual Multiple Errands Test (VMET)
Obsessive and Compulsive Disorders (OCD)	Virtual Multiple Errands Test (VMET)
Mood Disorder	Jansari assessment of Executive Functions Nesplora Aquarium

Table 1.1: Available VR-based tools for Psychiatric Disorders

Briefly, VR-based assessment tools for Schizophrenia allowed for detecting a deficit in performing shopping tasks due to impairments in multiple components of executive functionality, such as planning, mental flexibility, divided attention, and problem-solving (Aubin et al., 2018). Furthermore, VR-based tools appeared more sensible in detecting executive impairments than traditional neuropsychological tests (la Paglia et al., 2014). Finally, the ability of VR tasks to distinguish between schizophrenia subgroups based on the severity of executive dysfunction (mild vs severe) has a clear bearing on rehabilitation (Josman et al., 2009). In OCD assessment, VR-based assessment seems to be more sensitive to effects and changes in the executive sphere: patients made more mistakes, acted inefficiently, and scored higher in split attention and perseverance (persevering in errors is a clear sign of reduced flexibility) (la Paglia et al., 2014, 2016). Finally, preliminary findings on the application of VR assessment tools in mood disorders showed the ability of instruments to predict performance scores on neuropsychological tests indicating that they could use an index of EFs, particularly attention and inhibition (Hørlyck et al., 2021).

<u>Neurological diseases</u>: involve subjects with cognitive and functional impairments due to central nervous system dysfunctions (Elkind et al., 2001; Rizzo et al., 2000). The impairment of EFs is common in neurological patients and is related to brain dysfunction in the frontal cortex and its many interconnections with the posterior lobes and subcortical structures (Cipresso et al., 2014; la Paglia et al., 2014; Luria et al., 1966; Rizzo et al., 2000). For example, executive deficits are widely prevalent in Alzheimer's disease (over 75%) and stroke (between 40% and 75%) (Godefroy et al., 2018). To date, VR-based tools appear to be an encouraging

and interesting solution to the early evaluate executive impairments in a variety of neurological illnesses, which can be divided into three categories: acute, neurodegenerative, and neurodevelopmental (Table 1.2).

Neurological Diseases	
Acute Neurological Disease	
Acquired brain injury (ABI)	Jansari Assessment of Executive Function Virtual Classroom Multitasking in the City Test
Traumatic Brain Injury (TBI)	Avatar-Based Virtual Reality Virtual Library Task (VLT) Assessim Office
Post-Stroke	Virtual Action Planning Supermarket (VAP-S) Virtual Multiple Errands Test (VMET) Virtual Supermarket Adapted Four-Item Shopping Task
Neurodegenerative Disease	
Parkinson's Disease (PD)	Virtual Supermarket Virtual Multiple Errands Test (VMET)
Mild Cognitive Impairment (MCI) - Alzheimer's Disease (AD)	Virtual Reality Day-Out Task Virtual Action Planning Supermarket (VAP-S) Kitchen and cooking
Huntington's Disease (HD)	Virtual reality task – "EcoKitchen."
Multiple Sclerosis (MS)	Assessim Office (AO)
Neurodevelopmental Disorder	
Attention Deficit Hyperactivity Disorders (ADHD)	Continuous Performance Testing (based on Virtual Reality) Nesplora Aquarium Virtual classroom
Autism Spectrum Disorders	Virtual Classroom Bimodal Stroop task Virtual Errands Test

Table 1.2: Available VR-based tools for Neurological Disease

Briefly, executive dysfunctions in planning, prioritization, selection, creative and adaptive thinking, prospective memory, attention, cognitive flexibility, inhibition, and information processing speed were demonstrated in ABI using VR (Jansari et al., 2014). These findings were interesting because some authors showed that people with ABI normally performed on conventional clinical tests of EFs (Denmark et al., 2019), although showing impaired executive functioning in their daily life (Damasio, 1996). Importantly, VR made it possible to evaluate executive dysfunctions in children with ABI. Specifically, Virtual Classroom seemed to be a sensitive, playful, and ecologically valid assessment tool for diagnosing attention deficits (e.g., sustained attention) (Gilboa et al., 2019). In neurodegenerative disease, VR-based tools appeared to be valuable solutions in assessing executive dysfunctions since the initial stage of the disorder. In PD, for instance, VR-based tools allowed early evaluation of executive deficits (e.g., planning, decision-making, set-shifting), capturing impairments not detected by traditional neuropsychological assessment (Cipresso et al., 2014). Similarly, VR-based assessment appeared more sensitive to early executive deficits (e.g., planning cognitive flexibility, self-monitoring, and divided attention) in premanifest HD and manifest non-demented HD (Júlio et al., 2016, 2019), where patients presented a similar executive profile to healthy participants in the traditional neuropsychological battery. Over time, VR seemed to be a good, brief and user-friendly technique to test EFs in people with MCI and AD (at a very early AD phase) able to overcome the limitations (e.g., education, age, cultural background, other impairments, lack of ecological validity) of traditional tools examinations (Werner et al., 2009). Additionally, a VR-based tool exhibited great sensitivity and specificity as a screening test in discriminating MCI, mild AD and normal ageing (Tarnanas et al., 2013). Interestingly, the ability of assessment tools to collect executive impairments since the early stage of the disorder must consider a key feature in the neurodegenerative field as it could allow for identifying patients at risk of developing dementia and providing early neurorehabilitation interventions (Cipresso et al., 2014). Finally, VR can be considered a promising objective and reliable solution to evaluate deficits in executive functioning in neurodevelopment disorders. Interestingly, the integration between observation scales, objective cognitive profile and performance could allow specialists to achieve a more accurate and reliable assessment for the diagnosis of ADHD and provide specific recommendations for parents and teachers tailored to individual needs.

<u>Other clinical conditions</u>: a limited number of studies have shown promising preliminary results in using VR-based instruments to evaluate possible executive difficulties in particular conditions, for example, alcohol intoxication (planning, shifting, prioritization, creativity) or ecstasy (planning).

Other clinical Conditions	
Ecstacy/Alcohol Intoxication/Smoke/ Caffeine/Neurofribromathosis/Coronary Artery Disease	Virtual Classroom Bimodal Stroop task VR-cooking Task Jansari Assessment of Executive Function

Table 1.3: Available VR-based tools for Other clinical conditions

In conclusion, this summary amply shows that VR-based evaluation can be considered a new promising paradigm for neuropsychological assessment, able to provide an ecological evaluation of everyday executive impairments predicting real-world performance (Parsons, 2015), compared to traditional paper-and-pencil or computerized neuropsychological batteries (Neguț et al., 2016). To date, the applications of VR as an innovative solution for assessment are numerous, although many are still in the development stage and hence partially unavailable for use in clinical settings.

In recent years, one of the most trends in the VR technology field is the 360° technology (Negro Cousa et al., 2019) that appeared as an interesting instrument in the different healthcare sectors, including neuropsychological assessment (Negro Cousa et al., 2019; Pieri et al., 2021; Serino et al., 2017), rehabilitation (Bialkova & Dickhoff, 2019) and educational training (Violante et al., 2019).

1.6 Executive Functions and 360° Technology: A new opportunity?

Advances in 360° technology emerged as a valuable alternative approach to create VR content by recording familiar environments before, and then showing them to participants on a Head-Mounted Display (HMD) via smartphone (Negro Cousa et al., 2019; Ventura et al., 2019). In other words, 360° technology can be considered a promising interactive virtual technology for creating virtual-reality–immersive applications at a low cost (Violante et al., 2019).

Indeed, unlike the traditional VR settings, developing 360° environments does not require expensive specialized software or specific technological knowledge because they are made out of 360° photographs or videos that only need a standard 360° camera and free applications (Parsons & Phillips, 2016). The 360 ° technology can be incorporated in the "virtuality continuum" proposed by Milgram, in which the stimuli are presented in a space between the actual and the virtual, "a mixed reality", where extremes can coexist, producing new experiences (Milgram & Kishino, 1994). In this context, 360° technologies could allow participants to be evaluated in virtual environments that they experience from a first-person perspective without adverse clinical effects (e.g., nausea, vertigo), enhancing the global user experience of evaluation (Parsons, 2015). Indeed, preliminary studies involving 360°-based tools and neurological patients have shown the total absence of adverse events in 360° environments (Realdon et al., 2019; Serino et al., 2017). In this direction, Serino and colleagues have developed a 360° version of the Picture Interpretation Test (PIT) for the detection of executive deficits (only active visual searching component), successfully tested on patients with Parkinson's Disease and Multiple Sclerosis (Realdon et al., 2019; Serino et al., 2017) (Figure 1.6).

Figure 1.6: Screenshot of Picture Interpretation Test 360° (Serino et al., 2017)
Copyright: © The Author(s) 2017

Following these encouraging findings, we conceptualized, designed, and created the EXecutive-functions Innovative Tool 360° (EXIT 360°), a new 360°-based tool for achieving information about many components of executive functioning (e.g., planning, decision-making, problem-solving, attention, and working memory) (Borgnis, Baglio, Pedroli, Rossetto, Riva, et al., 2021). EXIT 360° aims to develop the PIT 360° by enabling the assessment of many components of executive functioning in ecological contexts.

This thesis aims to describe the EXIT 360° development process from its concept to validation as a sensitive tool for evaluating executive functionality. Overall, we will focus on the psychometric proprieties of EXIT 360°, analyzing its usability, convergent validity, and diagnostic validity, that is, its efficacy in discriminating between pathological and control groups. Parkinson's Disease (PD) was chosen as the clinical condition.

1.7 Why have we chosen Parkinson's Disease as a clinical population?

Parkinson's disease (PD) is a progressive neurodegenerative disease primary recognized as a movement disorder due to an extrapyramidal syndrome. Beyond the well-known motor symptoms (e.g., bradykinesia, resting tremor, rigidity, and postural instability), individuals with PD frequently experience a wide range of non-motor symptoms (NMS) since the start of the disease course, often even before the onset of motor symptoms in the prodromal state (Aarsland et al., 2017; Fang et al., 2020; Fengler et al., 2017). Cognitive impairment appears as one of the main clinical NMS that may be highly variable regarding the onset timing and the progression rate (Aarsland et al., 2005, 2007; Buter et al., 2008; Ceravolo et al., 2012; Hely et al., 2008; Kudlicka et al., 2011). In this context, executive dysfunction (ED) is the best-defined cognitive impairment in early-stage non-demented PD (Kudlicka et al., 2011). ED consists of deficits in attention, planning, set-shifting, dual-task performance, inhibitory control, working memory, and decision-making, even compromising social-cognition skills (Dahdah et al., 2017; Diamond, 2013; Dirnberger & Jahanshahi, 2013). As a result, patients have trouble in many essential goal-directed everyday activities, with critical adverse implications for daily functioning (i.e., preparing meals, managing money, shopping, and work) (Cipresso et al., 2014; Lawson et al., 2016; Levine et al., 2007) and quality of life (Barone et al., 2017; Lawson et al., 2016; Leroi et al., 2012). An increasing number of longitudinal studies have also revealed that early ED is a marker for PD conversion in "PD with dementia" (Azuma et al., 2003; Janvin et al., 2005). Therefore, early identifying executive deficits could permit identifying patients at risk of developing dementia, providing early neurorehabilitative interventions (Cipresso et al., 2014; Serino et al., 2014). As previously said, the ED was traditionally evaluated by paper-and-pencil neuropsychological tests that have proved unable to predict the real executive status since they don't reflect the complexity of EFs in everyday situations (Chan et al., 2008; Sirigu et al., 1995).

In this context, an early and ecologically valid assessment of everyday executive difficulties appears crucial to achieve excellent and effective disease management. Therefore, VR-based instruments that allow for carrying out different daily tasks in ecologically valid and controlled environments (Armstrong et al., 2013; Parsons, 2015; Serino et al., 2017) appeared to be a promising solution in the early evaluation of ED. Over the years, many authors have investigated ED in PD using VR-based tools that required participants to complete real-world tasks (such as shopping) in virtual settings that reproduced real supermarkets (Cipresso et al., 2014; Klinger et al., 2004, 2006).

A first study revealed deficits in planning and the switching mechanisms needed to manage a lot of information concurrently (Klinger et al., 2004, 2006). Clinicians assessed the strategic choices and planning skills of people with PD while purchasing specific products in a virtual supermarket. Inefficient use of contextual elements and a steady decline in spatial and temporal planning processes are suggested by the data, which demonstrated an increase in distance and duration and an inefficient trajectory in the PD group. Additionally, the inefficient trajectory suggested a dysfunction in the switching mechanism.

In the following years, the virtual version of MET made it possible to identify several common executive problems in PD patients (Cipresso et al., 2014; Raspelli et al., 2009). VMET enables the evaluation of patients' capacity to create and check a list of goals to successfully respond to environmental requests to complete several tasks (e.g., buy a specific product, or ask for information about a product to be purchased). In a preliminary study, Raspelli and colleagues confirmed the presence of deficits in planning and showed impairments in problem-solving, set-shifting, and sustained attention (few strategies and much perseveration) (Raspelli et al., 2009).

The next research conducted by Albani and colleagues revealed the presence of decision-making difficulties in individuals with PD. Patients made more errors and fewer effective strategies than controls, suggesting impulsive decision-making (Albani et al., 2010). Similar findings were made by Cipresso and colleagues, which supported cognitive flexibility impairments in PD with normal cognition (Cipresso et al., 2014). Patients performed worse on the VMET activities and used fewer efficient strategies to complete them. This study demonstrated that a real-world evaluation generates a more accurate assessment of the patient's impairment, which is hidden in conventional measures: patients with PD (PwPD) perform differently from healthy control subjects on the VMET but not on the traditional neuropsychological assessment of EFs (i.e., Clock Drawing Test, and Tower of London).

Therefore, a more ecologically valid evaluation of EFs leads to better detection of subtle deficits since the early stage of PD. It was an excellent result since early identification of executive impairments could help identify those individuals at risk of dementia and provide the opportunity for timely rehabilitation (Cipresso et al., 2014; Serino et al., 2014).

Recently, some authors have exploited 360° technology in evaluating executive functioning in PD (Serino et al., 2017). Specifically, Serino and colleagues have designed, developed and validated a 360° version of the Picture Interpretation Test (PIT), a paper-and-pencil test for EFs. The validation study has shown the effectiveness of PIT 360° as a highly sensitive ecological technique for identifying deficits in active visual perception since the early stages of PD. In addition, PIT 360° showed a good convergent validity, evaluated by the correlation with two traditional tests for EFs: Trail Making Test and the Phonemic Fluency Task. However, unlike conventional neuropsychological tests, PIT 360° was able to distinguish the pathological group from controls based on two indexes: time taken to arrive at the answer and the number of elements of the scene named. People with PD described the scene in great detail compared to healthy controls, took longer to provide the correct interpretation of the scene, and were more susceptible to interference from distractions. As a result, PD demonstrated greater difficulty focusing on the elements essential for the right scene evaluation.

As previously said, considering these interesting results, we have developed EXIT 360°. EXIT 360° wants to be an evolution of the PIT 360°, allowing for ecologically valid evaluation of more components of executive functioning. In this context, EXIT 360° could permit a quick, ecological, complete, and realistic assessment of ED of PD, allowing clinicians to tailor rehabilitative interventions according to patients' needs.

As stated earlier, this thesis will focus on the EXIT 360° development process from its concept to validation as a sensitive tool for evaluating executive functionality. In detail, Chapter 2 will thoroughly describe the concept, design, and development of EXIT 360°, highlighting the main innovative characteristics. In the following chapters, we will analyze the studies conducted to evaluate the main psychometric properties of the tool, respectively usability, convergent validity, and construct validity. The major findings of the usability studies involving PwPD and healthy control volunteers will be covered in Chapter 3. Chapter 4 will then detail the investigation findings to evaluate the convergent validity. The effectiveness of EXIT 360° in discriminating between pathological and control groups will then be examined in Chapter 5. Finally, we will discuss EXIT 360's potential as an innovative tool and potential future applications in light of the findings (Chapter 6).

2

2. EXIT 360° - From Concept to Validation

EXIT 360° was conceived as an innovative tool to provide a complete and integrated ecologically valid evaluation of executive functioning, involving participants in a "game for health" delivered via a common Head-Mounted Device (HMD). In detail, EXIT 360° was conceived and designed to be a brand-new task for EFs requiring subjects to carry out everyday subtasks in 360° environments that reproduce different real-life contexts. Participants can experience these 360° situations directly on HMD via a smartphone connected to the technological device.

The chapter will thoroughly describe the concept, design, and development of EXIT 360°, focusing on its key revolutionary features. Additionally, we'll provide a detailed explanation of the validation process used to assess the main psychometric properties of EXIT 360°.

2.1 CONCEPT, DESIGN AND DEVELOPMENT

2.1.1 EXIT 360°: Virtual Environments

Several evidence converges in supporting that 360° technology can be considered an interesting alternative approach to create VR content by recording familiar environments before, and then showing them to participants on an HMD (Negro Cousa et al., 2019; Ventura et al., 2019).

We have chosen to use 360° environments in designing EXIT 360° for three main reasons:

1. Developing 360° environments does not require expensive specialized software or technological competence since they consist of 360° photographs or videos that only need a standard 360° camera and free applications (Parsons & Phillips, 2016). For example, EXIT 360° environments were developed using a Ricoh Theta S Digital Camera, which enables to obtain 360° spherical imageries without high technical competencies and costs. The camera can

capture a 360° scene by combining two 180° scans via integrated software (resolution of 1792 × 3584 pixels) (Figure 2.1).

Figure 2.1: An example of 360° scenes used in EXIT 360° (here, the image is represented in anamorphic format)

2. 360° photos appeared in previous studies as a promising solution to provide an ecological evaluation of EFs (Serino et al., 2017). Clinicians can examine their patients while performing daily activities in ecologically controlled settings that reflect real scenarios and situations. In other words, 360°-based instruments offer real-world contexts, allowing patients to live a first-person experience "like in real life". This novel approach could allow evaluating how the patients face everyday tasks involving EFs, allowing a deeper comprehension of the patient's executive profile.

3. Previous studies involving 360°-based tools and neurological patients have shown promising results in terms of quality of user experience and usability. Firstly, studies have not demonstrated the presence of adverse events in 360° environments (Realdon et al., 2019; Serino et al., 2017). The lack of negative issues enhances the global user experience of evaluation since cybersickness could affect the performance and, consequently, the test's validity (Parsons, 2015). Moreover, thanks to the HMD, subjects experience the situations from a first-person perspective, exploring the entire space by moving their head without having to learn to use complex technological tools (e.g., joystick, mouse) (Serino & Repetto, 2018).

The existing VR-based tools for assessing EFs have always used everyday settings such as supermarkets, offices, library, and schools but fewer domestic settings (Borgnis, Baglio, Pedroli, Rossetto, Uccellatore, et al., 2022). On the contrary, the environments we have chosen for developing EXIT 360° reproduce everyday domestic contexts: a kitchen, two bedrooms, one living room and a landing. Due to the high impact of executive functioning in daily life, we believed that it would be beneficial to assess any executive impairments in the setting most experienced by the subjects, also in order to plan a rehabilitative intervention. EXIT 360° is also significantly different from other VR tools, where participants were immersed in a single scenario (e.g., virtual supermarket) and carried out only one task (e.g., shopping) (Borgnis, Baglio, Pedroli, Rossetto, Uccellatore, et al., 2022). Indeed, in the five domestic environments, participants must complete seven subtasks that correspond to daily life assignments, according to the environment in which they are immersed. The subtasks were conceived and designed to tap different EFs and were easily performed by projecting appropriate stimuli into the virtual domestic environments. As stimuli, we used standard 2D or 3D objects from free databases (for example, Paint3D – Figure 2.2)

Figure 2.2: An example of 2D and 3D objects implemented as stimuli in 360° environments

The InstaVR software, which permits to organize virtual environments and tasks in a unique experience, allowed for the integration between 360° settings and tasks.

2.1.2 EXIT 360°: Seven Subtasks

The participants' primary goal is to EXIT from a household environment in the shortest possible time by completing seven subtasks of increasing difficulty in 360° domestic immersive environments (Table 2.1).

	Name	**Place**
Task 1	Let’s Start	Neutral room
Task 2	Unlock the Door	Landing
Task 3	Choose the Person	Living room
Task 4	Turn On the Light	Corridor
Task 5	Where Are the Objects?	Bedroom
Task 6	Solve the Rebus	Bedroom
Task 7	Create the Sequence	Kitchen

Table 2.1: Subtasks and related 360° domestic environments

The seven subtasks were created to measure multiple dimensions of executive functioning swiftly and simultaneously. Each subtask assesses one or more EFs, including attention, working memory, planning, decision-making, and problem-solving. Depending on the cognitive load and the existence of confounding variables, the subtasks' level of complexity varies. The subtasks simulate typical everyday situations and ask the subject to complete assignments in accordance with their demands. The subjects must select from three or more "options" the one that will enable them to complete the task in the best way and move on with their journey. In the HMD, they see a small white dot, a "pointer" that follows their gaze. When subjects want to respond inside the environment, they simply need to move their heads and put the white dot over the response for a few seconds, and the answer will be selected automatically (Figure 2.3).

Figure 2.3: The representation of the small white dot that participants see in the headset.

This solution makes it possible to gather participant responses without becoming proficient with complicated tools (e.g., joysticks). The psychologist accompanies the participants along the entire path. However, all instructions are given in a standardized way to avoid the performance being adversely impacted by the various administration techniques (Rizzo et al., 2001). Specifically, the instructions were recorded during development and added to the virtual scenarios.

Description of Seven Subtasks and 360° Environments

Task 1: “Let’s Start”: In a neutral living room, participants must examine a map (Figure 2.4) and select the path from four choices that will lead them to the exit in the least amount of time. The subjects will select option "2" if they plan a correct strategy.

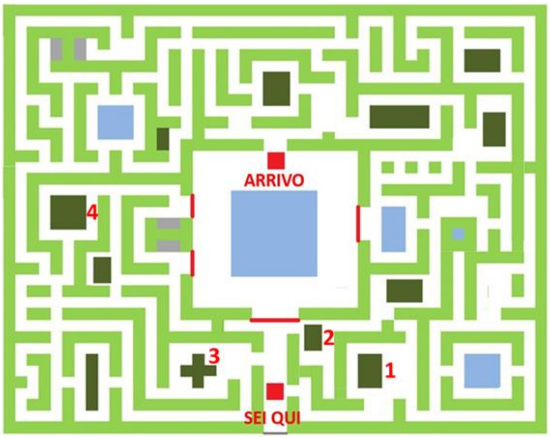


Figure 2.4: Task 1 - "Let's Start": Map with four alternative paths

Task 2 – "Unlock the Door": Participants are immersed in a 360° landing with a closed door. The recorded voice instructs the participants that they must first open the door to proceed and enter the house. Three 3D objects - a key, a drill, and a phone - appear in the hallway as the subjects investigate the place, and they must decide which of the three will open the door the best.

Task 3 – "Choose the Person": Participants were placed in a living room with five other people: 1) a man reading a book while sitting on a chair; 2) a mother and son speaking while watching television on the sofa; 3) a woman writing in her diary while sitting at the table; and 4) a girl sitting next to her using a computer for working. Participants must explore the room and select the correct person according to a precise instruction "*One of these people has a clue that allows you to move forward on your path*", followed by a description of the person to find. The subject will be able to choose the right person (the woman who is writing in her diary) only by concentrating on the entire description.

Task 4 - "Turn On the Light": The participants walk into a lengthy hallway lined with objects (e.g., table, pictures, and ornaments). After a few seconds, the corridor becomes completely dark. The participants heard the following instruction: "*It's all dark! The power went out. Here are four objects* (i.e., torch, unlit candle, ball, and lamp) *that could help you: you can only choose one*". The chamber will once more light up if the test subjects respond with the right response (torch): "*The light is back! We can continue with our journey*".

Task 5 - “Where Are the Objects?”: Participants are placed in a bedroom with a variety of furnishings, including a bed, chair, and dresser on which have different items strewn about them (e.g., candles, toys, lamps, glasses, and telephones) (Figure 2.5). Participants must identify the piece of furniture on which four distinct objects - a lamp, a plush toy, a blanket, and a phone - are arranged in a group. There are objects scattered throughout, but only the dresser has all of them.

Figure 2.5: 360° bedroom of Task 5 (here, the image is represented in anamorphic format).

Task 6 - “Solve the Rebus”: In a different bedroom, the participants must solve a rebus that appears on the wall and contains numbers and geometric forms in various colors (Figure 2.6). Subjects must comprehend the missing number and geometric shape. Then, they must select the correct answer from a list of options that display close to the rebus.

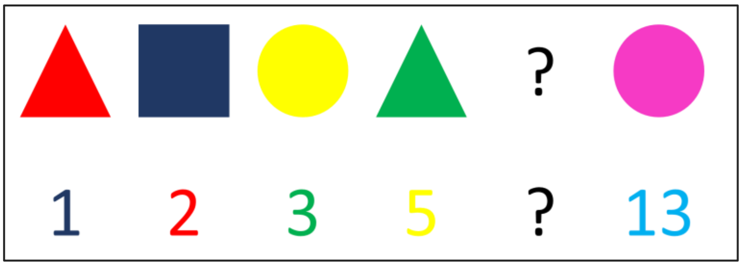


Figure 2.6: Task 6 “Solve the Rebus”

Task 7 - "Create the Sequence": Participants see an escape door in a kitchen that needs to be opened with a numerical combination. On top of a black sofa, a set of five numbers materializes in the space, one after the other (Figure 2.7).

Figure 2.7: Representation of Task 7: 360° kitchen showing one of the series numbers

The recorded voice says to the subjects: *"I ask you to read the numbers and try to memorize them. Our goal is to unlock the door to exit. The combination is formed by the series of numbers you will see but backwards, from last to first. If you see 2 3 5, you will have to say 5 3 2. What is the correct combination?*".

If the sequence reported is correct, the participants will see a map like the starting one which will contain the phrase "GOOD, NOW YOU ARE HERE!" and the completed path with all visited rooms marked (Figure 2.8).

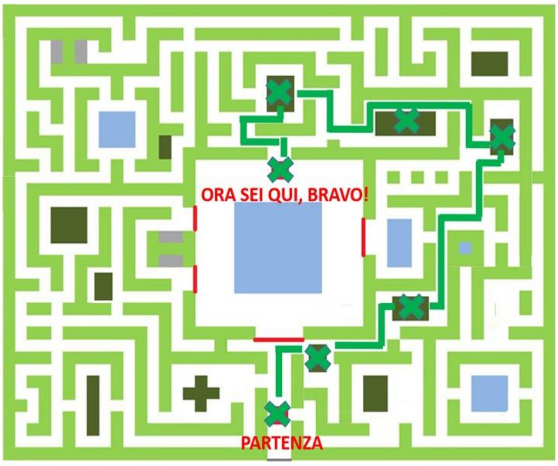


Figure 2.8: Final Map

2.1.3 EXIT 360°: Evaluation Procedure

Participants undergo an evaluation of about 15 min with EXIT 360°.

The psychologist started the administration by inviting the participant to sit on a swivel chair and wear the HMD. The psychologist gave the following general instruction on the task before the HMD was worn: "*You will now wear a headset. Inside this viewer, you will see some 360° rooms of a house. To visualize the whole environment, I ask you to turn on yourself; you are sitting on a swivel chair for this reason. Within these environments, you will be asked to perform some tasks*." Participants with presbyopia might put on their glasses.

A brief (1 minute) familiarization phase was used to introduce participants to the technology and observe potential adverse effects (e.g., dizziness, nausea). In this phase, participants were immersed in a neutral 360° environment resembling a living room, replete with a table with a plant, a sofa, several seats, and ornaments scattered about the space (Figure 2.9).

Figure 2.9: Neutral 360° environment that represents a living room

Participants were instructed by the examiner to freely explore the environment, describing what they noticed. After the familiarization phase, individuals were asked if they might have felt queasy or unwell. The test must be stopped right away by the examiner if any adverse effects appear. On the contrary, subjects were immersed in 360° scenarios that simulated a neutral living room. The experimental session started with a specific instruction: "*You are about to enter a house. Your goal is to get out of this house in the shortest time possible. To exit, you will have to complete a path and a series of tasks that you will encounter along your way. Are you ready to start?"*. During the session, subjects had to freely explore the 360° settings by moving their head like they would in a natural situation (Serino & Repetto, 2018). Moreover, as previously said, participants had to complete all seven subtasks.

2.1.4 EXIT 360°: Outcome measures

All participants completed all seven subtasks, and the psychologist recorded all their responses. If subjects selected an incorrect alternative, they received only one point (vs two points for a correct answer). Overall, the following indices were calculated:

1) Correct answer for each subtask and EXIT 360° Total Score (ranged between 7-14).

2) Subtask reaction time: time (in seconds) from the start of a subtask until the participant provided an answer.

3) EXIT 360° Total Time: time in seconds registered from the first instruction until the participant provided the last answer.

2.2 VALIDATION PROCEDURE

After developing EXIT 360°, we have conducted several studies to evaluate its main psychometric properties: usability, convergent validity, and construct validity. Before starting the experiment, a study protocol was drawn up (Borgnis, Baglio, Pedroli, Rossetto, Meloni, et al., 2021). The protocol study involved four steps performed in a one-session evaluation: (a) introduction phase; (b) pre-task evaluation; (c) EXIT 360° session; and (d) post-task evaluation (Figure 2.10).

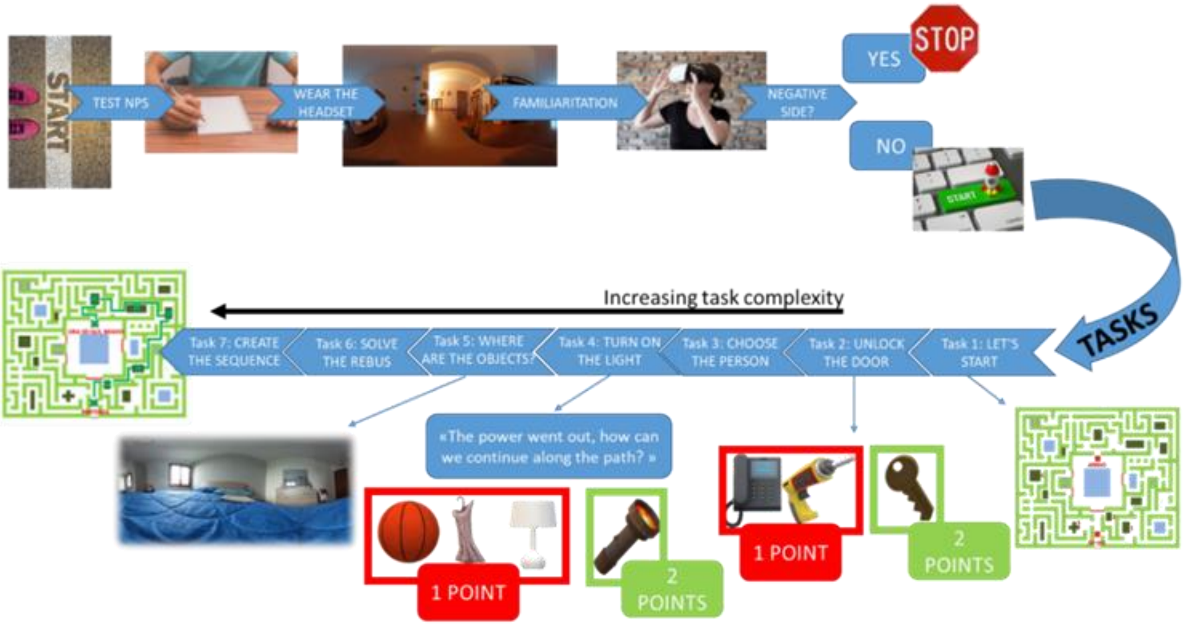


Figure 2.10: Evaluation process

The "Fondazione Don Carlo Gnocchi" and “Università Cattolica del Sacro Cuore” (Milan) Ethics Committees approved the study on April 2021. The psychologist fully explained the study's goals and risks to participants before they signed a written informed consent based on the revised Declaration of Helsinki (2013).

(a). *Introduction Phase*

After signing written informed consent, a psychologist gathered participants' socio-demographic data (e.g., age, gender, education level) and technological expertise using an ad-hoc questionnaire. The participants assessed their perceived familiarity and competence with various technologies, including tablets, smartphones, computers, and Internet (including social networks). The psychologist used a 5-point scale (ranging from "never" to "every day") to measure the familiarity with technology ("how often, in the last year, did you use..."), while a

5-point scale (ranging from "nothing" to "plenty") was used to assess their level of competence when utilizing those technologies ("how competent do you feel in using...").

(b) Pre-task Evaluation

Participants underwent a conventional paper-and-pencil neuropsychological assessment to obtain their global cognitive profile, focusing mainly on executive functioning. This evaluation allowed for compliance with the inclusion criteria (according to the aim of the study) and the convergent validity (comparing traditional neuropsychological tests and the EXIT 360° scores). The global cognitive level was evaluated with the Montreal Cognitive Assessment (MoCA), a sensitive screening tool to exclude the presence of cognitive impairment (Nasreddine et al., 2005; Santangelo et al., 2015). As regards executive profile, participants completed a complex battery including the Trail Making Test (in two specific sub-tests: TMT-A and TMT-B) (Reitan, 1992a, 1992b), phonemic verbal fluency task (F.A.S.) (Novelli et al., 1986), Stroop Test (Stroop, 1935), Digit Span Backward (Monaco et al., 2013), Frontal Assessment Battery (FAB) (Appollonio et al., 2005; Dubois et al., 2000), Attentive Matrices (Spinnler & Tognoni, 1987a) and Progressive Matrices of Raven (Caffarra et al., 2003; Raven, 1947). Table 1 describes the EFs evaluated by each neuropsychological test mentioned before

Name	Executive Function
Trail Making Test	Visual attention Task switching
Verbal fluency	Access to vocabulary on phonemic key
Stroop Test	Inhibition
Digit Span Backward	Working memory
Frontal Assessment Battery	Abstraction Cognitive flexibility Motor programming/planning Interference sensitivity Inhibition control
Attentive Matrices	Visual search Selective Attention
Progressive Matrices of Raven	Sustained and selective attention Reasoning

Table 2.2: Traditional neuropsychological tests and corresponding evaluated executive functions

(c). EXIT 360° Session:

After the initial screening, all participants underwent an evaluation with EXIT 360° (see above for EXIT 360° evaluation procedure).

(d) Post-Task Evaluation

After EXIT 360° session, subjects had to rate the usability and quality of the user experience of EXIT 360° using standardized scales and questionnaires.

The usability was evaluated using the System Usability Scale (SUS) (Brooke, 1986, 1996), a brief questionnaire with ten items on a 5-point scale ranging from "completely disagree" to "strongly agree". SUS is a rapid, efficient, reliable scale commonly used to assess the usability of a wide range of technological products. The total score, which ranges from 0 to 100, represents the global system usability. Additionally, the scale accounts for two key factors that influence the user experience: usability, which measures how easily a user can use a system (scores 1-4), and learnability, which measures how easily a user can learn to use it (scores 1-4) (Lewis & Sauro, 2009).

The user experience quality was assessed by:
(a) Flow Short Scale: three items (5-point scale: from low to high) to assess the perceived levels of task-coping skills, challenges, and the challenge-skill balance (Engeser et al., 2005).
(b) "Enjoyment" subscale of the Intrinsic Motivation Inventory (IMI) (Deci et al., 1994): four items (5-point scale: from low to high) to gauge participants' enjoyment of the suggested activity (i.e., boring, pleasant, fun and activating). The scores of item "boring" were reversed to align with the remaining items; therefore, on the whole scale, a low value reflects a negative perception of the experience with EXIT 360°.
(c) Sense of Presence Inventory (ICT-SOPI) scale: 44 items (5-point scale: from 1: "strongly disagree" to 5: "strongly agree") divided into two parts in order to evaluate thoughts and feelings after (Part A) or while (Part B) the user was experiencing the environment (Part B) (Lessiter et al., 2001). The 44 items allow for evaluating four subgroups, generated by calculating a mean of all completed items contributing to each factor:

(1) "Spatial physical presence" (19 items): feeling of occupying and controlling a physical space in the virtual environment (e.g., "*I felt I could interact with the environment shown*");
(2) "Engagement" (13 items): propensity to feel pleasantly and psychologically immersed in the virtual setting ("*I was sorry my experience was over*");
(3) "Ecological validity" (6 items): the tendency to consider the virtual environment as real (e.g., "*The environment shown seemed natural to me*");
(4) "Negative effects" (6 items): adverse psychological reactions (e.g., "*I felt nauseous*").

(d) User Experience Questionnaire (UEQ): a 26-item semantic differential scale: each item consists of two opposite adjectives (i.e., boring vs exciting) (Laugwitz et al., 2008; Schrepp et al., 2014, 2017; Schrepp & Thomaschewski, 2019). Six domains can be computed using the 26 items:

(1) Attractiveness: global impression of the tool
(2) Perspicuity: easily to learn to use the product
(3) Efficiency: users' effort to complete tasks
(4) Dependability: the feeling of interaction's control
(5) Stimulation: motivation to use the product
(6) Novelty: innovation and creativity of tool

3

3. EXIT 360° - Usability Studies

A systematic review conducted by Borgnis and colleagues (2022) has shown that several authors support the feasibility, acceptability, and effectiveness of VR-based tools in the early assessment of executive dysfunction in psychiatric and neurologic diseases (Aubin et al., 2018; Camacho-Conde & Climent, 2020; Cipresso et al., 2014; Dahdah et al., 2017). The use of 360° VR video for an ecologically valid evaluation of EFs in the neurologic population has recently attracted increasing attention (Realdon et al., 2019; Serino et al., 2017). In this framework, we developed EXIT 360°, conceived and designed to be an original 360°-based instrument for an ecologically valid evaluation of multiple components of executive functioning (Borgnis, Baglio, Pedroli, Rossetto, Riva, et al., 2021). After developing EXIT 360°, we conducted several studies to evaluate its main psychometric proprieties, including usability, convergent, and construct validity. In this Chapter, we will focus on the usability dimension in healthy control subjects and a clinical population, namely Parkinson's Disease.

The crucial importance of evaluating usability and quality of user experience when designing, developing, and using VR-based instruments has been demonstrated in a wide number of studies (Pedroli et al., 2013, 2018). Recently, Sauer and collaborators have proposed an innovative higher-level concept, the "interaction experience", according to which integrating these two critical aspects will provide greater benefits to users and improve their experience with technological tools (Sauer et al., 2020).

The usability assessment allows for knowing and comprehending the "degree to which a subject is able to use a system to achieve specific goals effectively, efficiently, and satisfactorily within a well-defined context of use" (Iso, 1998). In more detail, usability is made up of three primary components: 1) effectiveness (i.e., the probability for users to reach goals), 2) efficiency (i.e., users' efforts to achieve the goal), and 3) satisfaction (i.e., users' perceptions of their interactions with the product). Overall, the usability assessment enables comprehension of any technical issues that can impair subjects' performance.

The user experience is instead described as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service" (ISO & Standard, 2010). Over time, researchers have worked on five primary domains in the creation of digital contents to improve the user experience: “sense of presence”, “sense of realism”, “engagement”, “enjoyment”, and “side effects” (Aubin et al., 2018; Schultheis & Rizzo, 2001). Firstly, the developers have tried to increase engagement in virtual environments by improving the ecological validity and sense of presence and realism of VR-based tools. Specifically, the VR instruments want to offer scenarios comparable to the real world, allowing patients to live a first-person experience “like in real life”. Moreover, focusing on enjoyment and attractiveness increases users' motivation and participation while lowering anxiety linked to neuropsychological testing. Additionally, previous evidence showed sickness in virtual settings can include nausea, headaches, dizziness, and sweating (Armstrong et al., 2013). Users who have cybersickness may have unpleasant experiences that impact their performance and greatly reduce the validity of test outcomes. For example, headaches are a frequent symptom in the population with Traumatic Brain Injury; consequently, a tool that could exacerbate these symptoms would affect the performance, reducing the validity of the test results. Finally, evaluating users' technology familiarity, particularly older adults, is important since bad test results may be caused by a lack of understanding of how VR works (Parsons & Phillips, 2016). For all these reasons, accurate work on usability and user experience dimensions appears crucial.

In this Chapter, we will show whether EXIT 360° can satisfy all of these significant characteristics: good usability, absence of side effects, high sense of presence, ecological validity, enjoyment, and attractiveness. For this purpose, we have conducted two studies involving healthy controls (Borgnis, Baglio, Pedroli, Rossetto, Isernia, et al., 2021) and patients with Parkinson’s Disease (Borgnis, Baglio, Pedroli, Rossetto, Meloni, et al., 2022).

Study 1: EXIT 360° Usability and User Experience in Healthy Control Subjects

A systematic review highlighted that there are few studies on usability and user experience, despite their apparent importance in creating virtual tools. Only three research on 80 papers analyzed focused on usability evaluation, and their findings were inconsistent (Borgnis, Baglio, Pedroli, Rossetto, Uccellatore, et al., 2022). In Study 1, we evaluated the usability and user experience quality of EXIT 360° involving seventy-six healthy control subjects.

Material and Methods

Participants: 76 healthy volunteers were recruited at IRCCS Fondazione Don Carlo Gnocchi in Milan, based on the inclusion criteria: (a) age ranged 18-90 years old; (b) education level ≥5; (c) no cognitive impairment evaluated using Montreal Cognitive Assessment test (Nasreddine et al., 2005) (MoCA score ≥17.54, cut-off of normality), corrected for age and education level according to Italian normative data (Santangelo et al., 2015); and (d) ability to provide written, signed informed consent. Exclusion criteria were: (a) visual or hearing impairment that might have hampered the evaluation with EXIT 360°; (b) systemic, psychiatric, or neurological diseases; and (c) overt visual hallucinations or vertigo.

Procedures: Participants underwent an evaluation that involved three of four phases described in section EXIT 360° Procedure of study. Briefly, participants underwent a one-session evaluation involving: 1) technological expertise assessment (Introduction); 2) EXIT 360° Session and 3) usability and user experience evaluation using SUS scale, ICT-Sense of Presence and UX questionnaires (Post-Task Evaluation).

Statistical Analysis: Descriptive statistics included frequencies, percentages, and median and interquartile range (IQR) for categorical variables and mean and standard deviation (SD) for continuous measures. The Kolmogorov-Smirnov test was used to determine the normality of the data distribution. The scores of usability, user experience, and technological experience were compared using Pearson's correlation. Additionally, a one-way ANOVA non-parametric (Kruskal–Wallis) and Chi-square were calculated to confirm any potential differences in education and gender in the age group. Furthermore, a one-way ANOVA between subjects (post-hoc: Bonferroni test) was conducted to evaluate any significant differences in technological expertise due to age group. Finally, ANCOVA between subjects (covariate: education) was performed to examine possible differences in usability scores due to age. Jamovi

1.6.7 software was used to conduct all statistical analyses. A statistical threshold of $p < 0.05$ was considered statistically significant.

Results

Participants: Table 3.1 reports the demographic and clinical characteristics of the entire sample. Subjects (n = 76) were primarily female (M:F = 28:48) with a mean age of 53.5 years (SD = 20.30, range = 20–89) and education age nearly 13 (IQR = 13–18, range 5–18). All subjects showed no cognitive impairment (MoCA_adjusted score = 25.9 ± 2.63).

		Subjects (*n* = 76)
Age (years)	*Mean (SD)*	53.5 (20.30)
Sex (M:F)		28:48
Age of education (years)	*Median (IQR)*	13 (13–18)
MoCA_raw score	*Mean (SD)*	26.9 (2.39)
MoCA_adjusted score	*Mean (SD)*	25.9 (2.63)

Table 3.1: Demographic and clinical characteristics of the whole sample. M = male; F = female; SD = standard deviation; IQR = interquartile range; n = number; MoCA = Montreal Cognitive Assessment.

The demographic scores of the participants, divided by age group, are shown in Table 3.2. A significant correlation appears between the seven age groups in education level (χ^2 (6) = 29; $p < 0.001$), but not in the sex variable (χ^2 (6) =7.1; $p = 0.312$).

	20–29 (*n* = 13)	**30–39 (*n* = 11)**	**40–49 (*n* = 10)**	**50–59 (*n* = 11)**	**60–69 (*n* = 11)**	**70–79 (*n* = 10)**	**80–89 (*n* = 10)**	**Group Comparison *p-Value (*)***
***n* (%)**	17.1%	14.5%	13.2%	14.5%	14.5%	13.2%	13.2%	
Sex (M:F)	4:9	3:8	6:4	2:9	4:7	6:4	3:7	0.312
Ed. Median (IQR)	16 (16–18)	18 (17–18)	16.5 (13–18)	13 (13–13)	8 (8–13)	10.5 (8–13)	13 (5.75–16.8)	< 0.001*

Table 3.2: Demographic characteristics of participants divided by the age group. M = male; F = female; Ed. = Age of education SD = standard deviation; IQR = interquartile range; n = number; (*) significant difference

Specifically, results indicated a significant difference between group 20–29 and respectively, 60–69 (W = −4.711; $p < 0.05$) and 70–79 (W = −4.711; $p < 0.05$). Furthermore, a significant difference appeared between group 30–39 and respectively 50–59 (W = −4.31; $p < 0.05$), 60–69 (W = −5.04; $p < 0.05$), and 70–79 (W = −4.82; $p < 0.05$).

Technological Expertise: The mean score of the ad hoc 5-point questionnaire for measuring perceived technological familiarity was 3.41 ± 1.8, that is, "participants used the technology about once a week". In detail, 25% of individuals indicated a low (<3) familiarity with technologies, while 28.9% demonstrated a good (4—at least twice or three times per week) level of knowledge. Additionally, the ad hoc 5-point competency questionnaire had a mean score of 3.48 ± 1.11, which falls between "little" and "enough". Specifically, 28.9% of participants demonstrated poor (<3) technological competence, and 30.03% revealed strong (≥4—enough or much) competence across a range of technologies. Data revealed a significant difference between age groups both in familiarity (F (6,69) = 12.6; p<0.001) and competence (F (6,69) = 22.2; p<0.001) with technologies. In detail, results revealed a significant difference in familiarity between the 80-89 age group and every other age group except 70-79 (p > 0.05). Additionally, there was a noticeable difference between groups 70-79 and, respectively, 20-29 (p<0.001) and 30-39 (p<0.001). Regarding competence level, results showed a significant difference in the degree of competence between group 80-89 and all other age groups (p<0.05). Furthermore, there was a significant difference between groups 70-79 and 20-29, 30-39, and 40-49 (p<0.001, p<0.001, and p<0.05, respectively). Also, group 60-69 demonstrated a substantial difference with groups 20-29 (p<0.001) and 30-39 (p<0.05). Finally, only group 20-29 differed from group 50-59 (p<0.05).

Usability: The mean value of the global usability, calculated using the SUS, was 75.9 ± 12.8, suggesting a satisfactory degree of usability, according to the scale's score acceptability ranges (cut off = 68) and adjective ratings (Figure 3.1). Specifically, more than 77% of participants had scores higher than the cut-off.

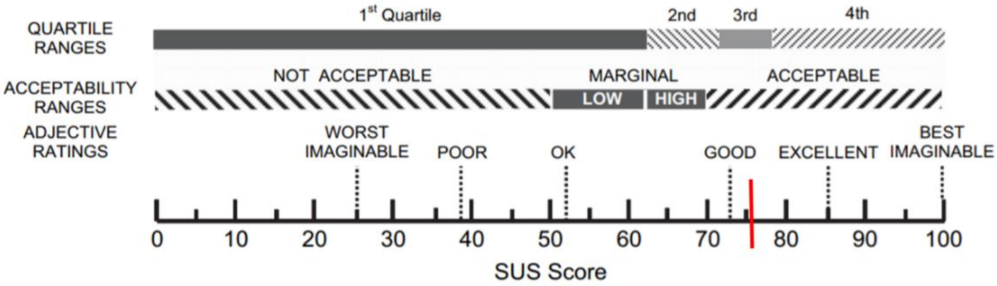


Figure 3.1: A graphic representation of the SUS score.

Additionally, 36.8% of participants evaluated EXIT 360° as "good", 32.9% as "excellent", and 27.6% as "best imaginable" according to the adjective rating (Figure 3.2).

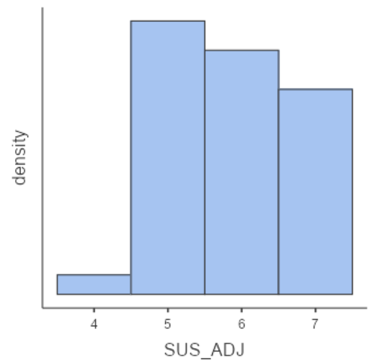


Figure 3.2: Graphic representation of the adjective rating.

Interestingly, participants provided good and promising ratings for learnability (mean = 2.91 ± 0.738) and usability (mean = 3.07 ± 0.55), two key factors that influence the user experience. Only 14.5% and 15.8% of participants showed low scores (<2.5) respectively at usability (easy-to-use) and learnability (easy-to-learn).

User Experience: The three items of the Flow Short Scale provided a high score for perceived skill in completing EXIT 360° (median = 5, IQR = 4-5) and enabled users to assess the challenge level of EXIT 360° based on their own abilities, such as "balance/appropriate" (median = 3; IQR = 3). Figure 3.3 shows that the "Enjoyment" subscale of IMI received high scores (≥4) in all four items: 1) boring (median = 5, IQR = 4–5), 2) enjoyable (median = 4, IQR = 3–4), 3) activating (median = 5, IQR = 4–5), and 4) funny (median = 4, IQR = 4–4.25).

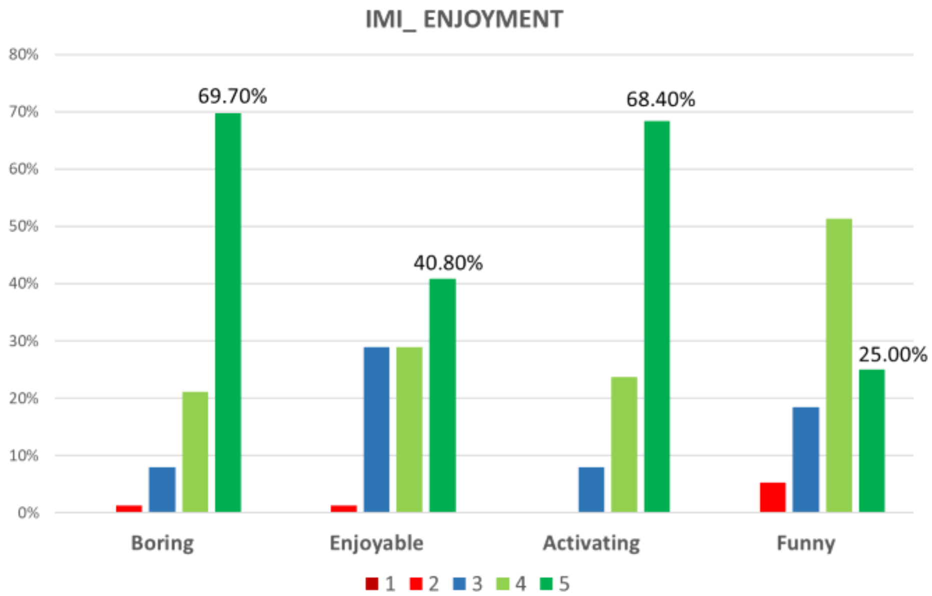


Figure 3.3: Graphic representation of the domains of the Intrinsic Motivation Inventory.

Figure 3.4 shows high scores in all ICT-SOPI dimensions: a) spatial presence (mean = 3.38 ± 0.55, range = 2.11–4.63), b) engagement (mean = 3.76 ± 0.56, range = 2.38–4.77), c) ecological validity (mean = 4.32 ± 0.54, range = 3–5), and d) negative effects (mean = 1.79 ± 0.95, range = 1–4). In detail, 10.4% (n = 11) of participants indicated the presence of adverse effects (such as nausea and dizziness), but only one person really displayed relevant side effects. Additionally, 75% demonstrated strong spatial presence (only 9.2% demonstrated scores below 2.5 - "*I felt I could interact with the environment shown*"), and 89.5% showed strong engagement when completing the EXIT 360° (e.g., "*I would have liked the experience to continue*"). Finally, 97.4% of people provided a good score in EXIT 360° ecological validity ("*I had the feeling that the environment shown was part of the real world*"), with more than 84% displaying high scores (≥4).

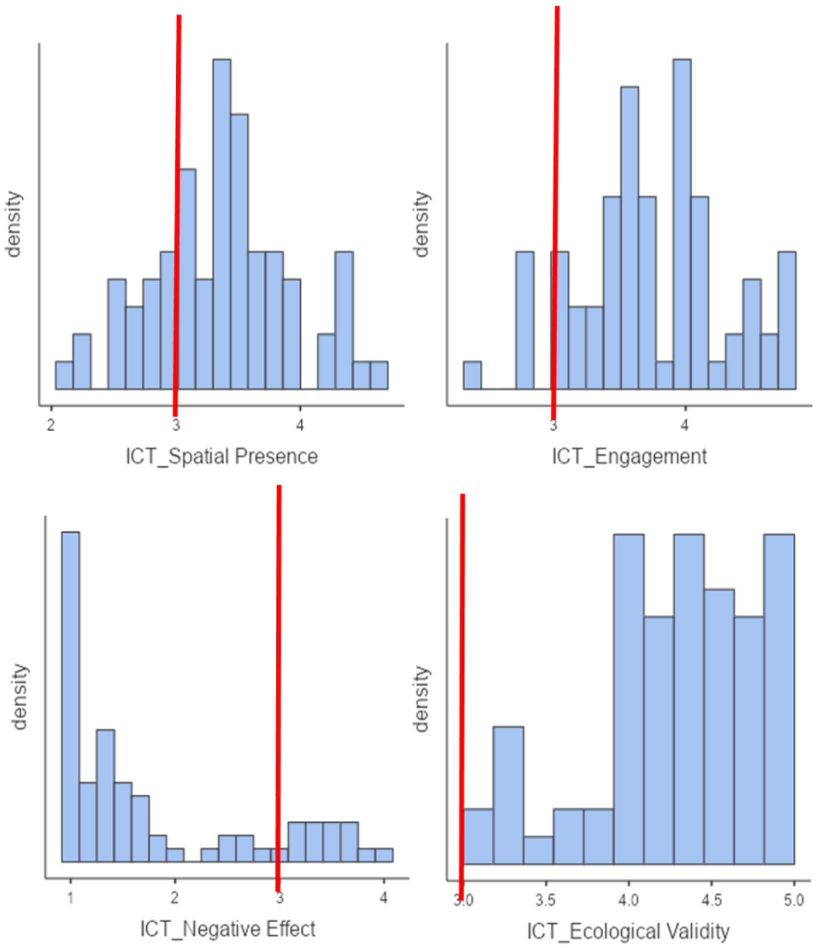


Figure 3. 4: Graphic representation of five ICT-SOPI dimensions. The red lines indicate a neutral score.

The UEQ questionnaire showed positive evaluation (>0.8) in all 26 items. Figure 3.5 shows good scores on all scales according to the questionnaire's ranges (between −3, horribly bad, and +3, extremely good).

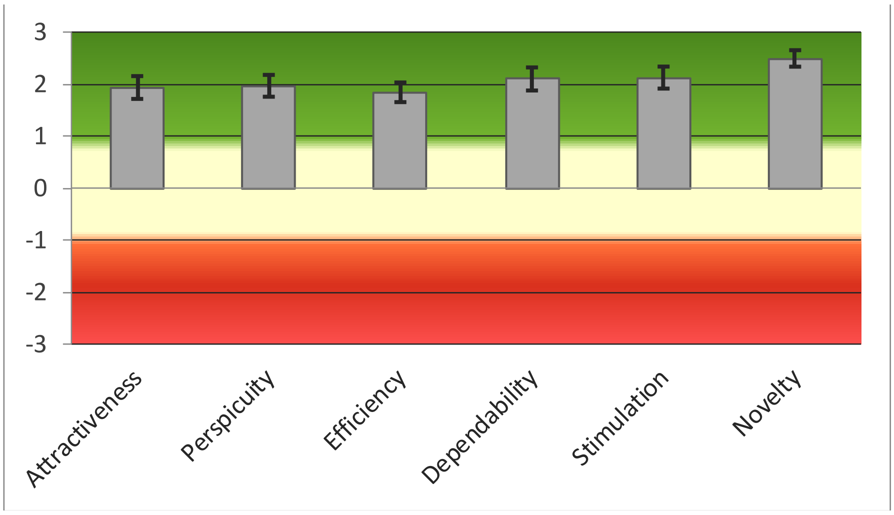


Figure 3.5: Graphic representation of scores of the six UEQ scales.

Table 3.3 details the mean scores of all scales together with their corresponding high internal consistency values (Alpha-coefficient > 0.7) (Cronbach, 1951).

	Mean	**SD**	**Confidence Interval**		**Alpha-Coefficient**
Attractiveness	1.934	0.976	1.715	1.715	0.92
Perspicuity	1.967	0.934	1.757	1.757	0.87
Efficiency	1.842	0.840	1.653	1.653	0.79
Dependability	2.099	0.986	1.877	1.877	0.86
Stimulation	2.125	0.942	1.913	1.913	0.88
Novelty	2.493	0.705	2.335	2.335	0.90

Table 3.3: Scores of the six UEQ scales. SD = standard deviation.

Additionally, the scales of the UEQ can be divided into two categories: hedonic quality, which includes non-task-related quality aspects like stimulation and originality, and pragmatic quality, which includes perspicuity, efficiency, and reliability. EXIT 360° specifically received a high rate for hedonic quality (mean = 2.31) and a good score for pragmatic quality (mean = 1.97). The means of each UEQ scale were then contrasted with preexisting values from a benchmark data collection (Schrepp et al., 2017) that included information from 20,190 participants across 452 research. The scales “attractiveness”, “dependability”, “stimulation”, and “novelty” received excellent evaluations, falling within the range of the 10% best results (Figure 3.6).

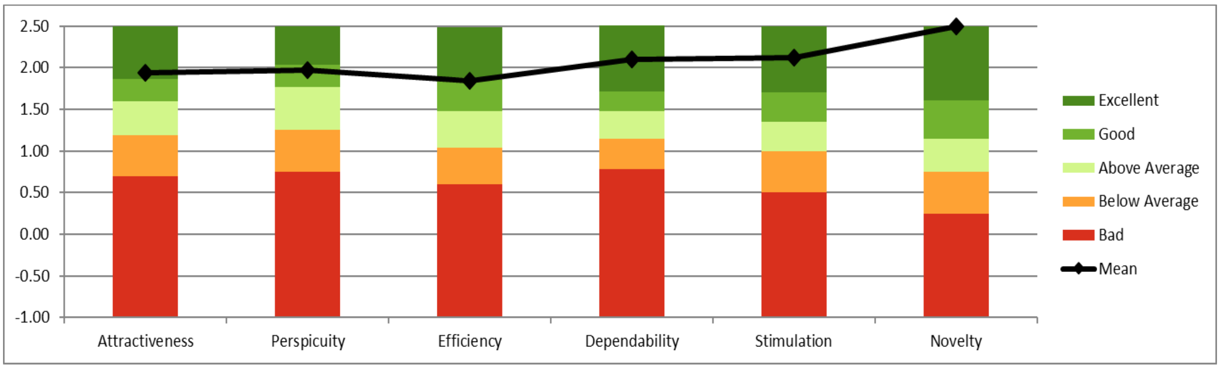


Figure 3.6: Comparison between means of each UEQ scale and values from a benchmark data set.

Correlation: Pearson's correlation showed no significant linear relationship between the SUS total score and the demographic factors, particularly education (r = −0.107; p = 0.356) and age (r = −0.078, p = 0.503). Furthermore, the ANCOVA-between subjects (covariate: education) revealed no significant differences in the SUS total score by age group (F (6,68) = 1.02; p = 0.419). It's interesting to note that this research also demonstrated that education level did not affect the SUS overall score (p = 0.159). A lack of a significant link was also shown between the SUS total score and technological expertise both in familiarity (r = 0.012; p = 0.915) and competence (r = 0.177; p = 0.127). In contrast, a significant positive linear correlation appeared between the SUS total score and three ICT-SOPI domains: spatial presence (r = 0.247; p<0.05), ecological validity (r = 0.466; p<0.001) and engagement (r = 0.495; p<0.001). Finally, Figure 3.7 shows a substantial and negative association (r = -0.43; p<0.001) between the ICT-SOPI domain "negative effect" and the SUS total score.

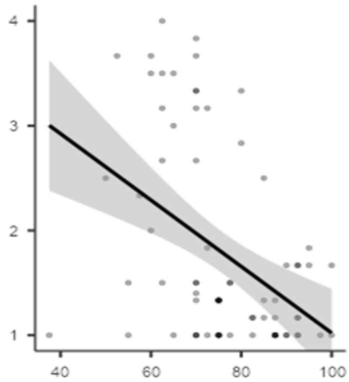


Figure 3.7: Negative correlation between adverse effects and SUS total score

Conclusion

The results of the present study are promising and interesting regarding the usability and user experience quality of EXIT 360° (Borgnis, Baglio, Pedroli, Rossetto, Isernia, et al., 2021). Firstly, all participants successfully carried out EXIT 360° without relevant adverse effects. Overall, healthy control subjects gave a positive global impression of the tool, describing it as usable, learn-to-use, appropriate to their abilities, clear, enjoyable, attractive (i.e., not boring, activating, funny) and friendly. Moreover, EXIT 360° was praised for being a quick, understandable, and effective tool with excellent hedonic quality in terms of stimulation (fun and engaging) and originality. Furthermore, EXIT 360° appeared to be a pleasure and challenging instrument with a strong spatial presence, outstanding ecological validity, and irrelevant negative effects (only one subject reported some adverse effects such as nausea and vertigo). Interestingly, neither technological nor demographic features had an impact on the encouraging outcomes. Finally, the overt relationship between usability and user experience quality supported the concept of interaction experience postulated by Sauer and colleagues, where usability and user experience simultaneously showed up as crucial components to give users more advantages while utilizing technological equipment (Sauer et al., 2020).

Study 2: EXIT 360° Usability and User Experience in Parkinson's Disease

To date, only one study has been conducted to test the usability of VR-based instruments for assessing executive abilities in Parkinson's disease, despite the evident value of paying attention to usability and user experience. In a pilot study, Pedroli and colleagues evaluated three PD patients and 21 healthy control subjects using the Virtual Multiple Errands Test (VMET). Data revealed that healthy participants awarded the VMET good usability, while patients with PD (PwPD) demonstrated that the VMET requires more than just an improvement to be deemed usable. Results also indicated that applying the virtual protocol to PD patients requires a strong training phase before the test (Pedroli et al., 2018).

In Study 2, we evaluated the usability and user experience quality of EXIT 360° involving 27 PwPD compared to 27 healthy control subjects (HC).

Material and Methods

Participants: 27 PwPD and 27 HC matched for age, sex (M:F = 11:16) and education were recruited at IRCCS Fondazione Don Carlo Gnocchi ONLUS in Milan. All participants had to meet inclusion criteria (a) age ranged 18-90 years old; (b) education level ≥5 (primary school); (c) no cognitive impairment as measured by MoCA, corrected for age and education following Italian normative data (Santangelo et al., 2015); and (d) ability to give a written, signed informed consent. Additionally, PwPD had to satisfy the following conditions for inclusion: (a) clinically established or probable Parkinson's disease in accordance with Movement Disorder Society (MDS) criteria (Postuma et al., 2015); (b) mild to moderate disease staging with scores < 3 on the Hoehn and Yahr scale; and (c) deficits in EFs verified by documented neurological and/or neuropsychological evaluation. Severe hearing or vision impairments, severe systemic, mental, or other neurological diseases, and overt visual hallucinations or vertigo were grounds for exclusion from the study.

Procedures: Following the procedure detailed in the "EXIT 360° Procedure of study" section, all participants underwent a one-session evaluation involving: 1) technological expertise assessment (Introduction); 2) A brief neuropsychological evaluation using the MoCA test and Frontal Assessment Battery (Pre-Task Evaluation); 3) EXIT 360° Session and 4) usability and user experience evaluation using SUS scale, ICT-Sense of Presence and UX questionnaires (Post-Task Evaluation).

Statistical Analysis: Descriptive statistics comprised frequencies, percentages, and median and interquartile range (IQR) for categorical measures and mean and standard deviation (SD) for continuous variables. The Kolmogorov-Smirnov test was used to determine the normality of the data distribution. The demographic and clinical characteristics and technological expertise of pathological and healthy groups were compared using a t-test for an independent sample (parametric or non-according to variables) and Chi-square test. Pearson's correlation was used to compare the usability ratings, user experiences, and technology experiences. A t-test for the independent sample was also performed to assess any significant variations in the same variables across the groups. Jamovi 1.6.7 software was used to conduct all statistical analyses. The statistical significance was defined as $p<0.05$.

Results

Participants: The demographic and clinical details of the entire sample (N = 54), divided into two groups, are shown in Table 3.4. PwPD (n = 27) was predominantly female (M:F = 11:16) with a mean age of 68.2 (SD = 9, range = 53–84) and age of education =13 (IQR = 5, range 5–18). HC was mostly female (M:F = 11:16) with a mean age of 66.4 (SD = 10.5, range = 48–88) and education = 13 (IQR = 5, range 5–18). No significant differences appear between the PwPD and HC in all major demographic and clinical variables. All participants demonstrated the absence of cognitive impairment (cutoff of normality= MoCA score ≥ 17.54).

	PwPD N = 27	**HC N = 27**	**Group Comparison (*p-Value*)**
Age (years, mean (SD))	68.2 (9)	66.4 (10.5)	0.507
Sex (M: F)	11:16	11:16	1.000
Age of education (years, median (IQR))	13 (5)	13 (5)	0.740
MoCA_adjusted score (mean (SD))	25.3 (2.25)	26.0 (2.53)	0.246

Table 3. 4: Demographic and clinical characteristics of the whole sample. M, male; F, female; SD, standard deviation; IQR, interquartile range; n, number; MoCA, Montreal Cognitive Assessment; PwPD, patients with Parkinson's disease; HC, healthy controls.

Technological Expertise: PwPD and HC provided similar rates on the ad hoc five-point scale for assessing perceived technology familiarity, with mean scores of 3.15 ± 0.89 and 3.14 ± 1.12, respectively (i.e., participants used the technology about once a week). Each group's percentages relating to familiarity are shown in Figure 3.8.

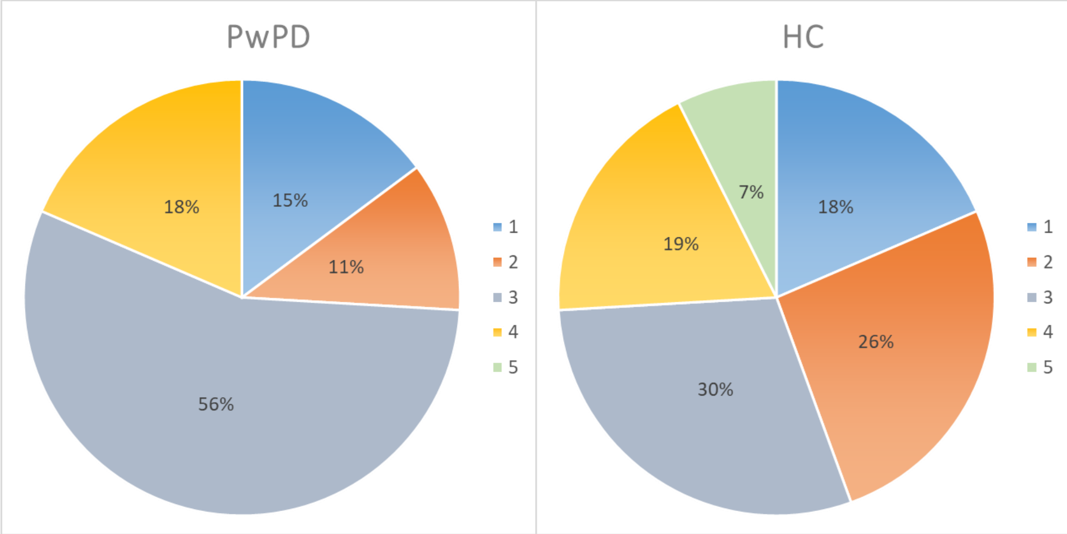


Figure 3.8: Percentages relating to familiarity with the technologies for each group. PD, Parkinson's disease; HC, healthy control. 1 = never; 2 = once a month or more rarely; 3 = once a week; 4 = every 2/3 days; 5 = every day.

Additionally, PwPD had a mean score of 2.68 ± 1.01 on the ad hoc five-point competency questionnaire, which is close to the "little" threshold. The average score for HC was 3.04 ± 0.98 (i.e., neither enough nor little). Figure 3.9 shows the percentages of each group's self-reported competence with various technologies. Only 7.4% of PwPD and 18.5% of HC demonstrated good (≥4—enough or much) technological competence.

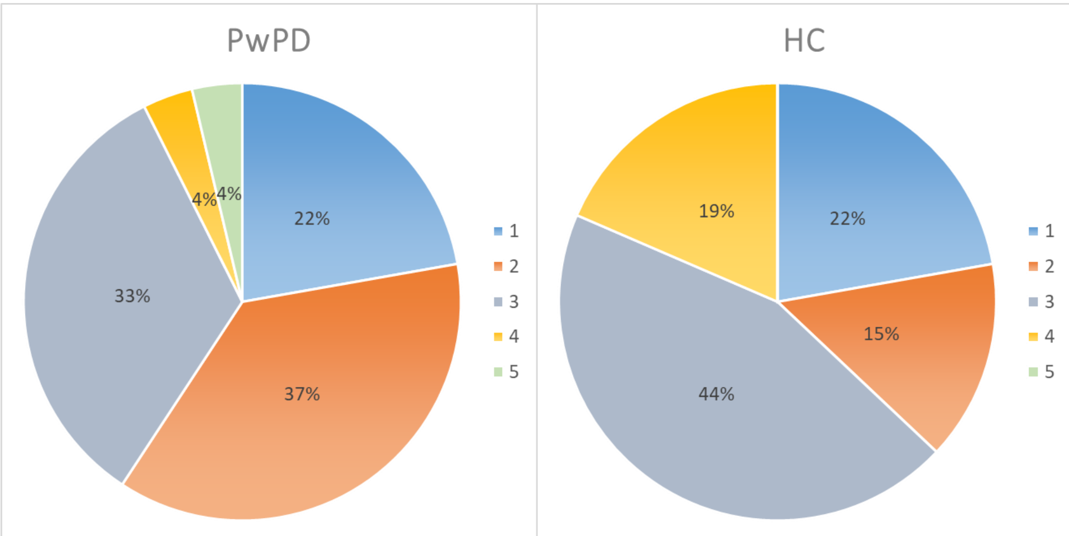


Figure 3.9: Percentages relating to self-reported competence in using several technologies for each group. PD, Parkinson's disease; HC, healthy control.

Analyzing the potential group differences, data revealed no significant differences in technology competence (t-test (52) = -1.377; p = 0.174) or familiarity (t-test (52) = 0.045; p = 0.964) levels.

Neuropsychological Evaluation: Preliminary analyses suggested that EXIT 360° may be a useful technique for discriminating PwPD from HC. In detail, Table 3 shows significant differences between the two groups in Total EXIT score (t-test (52) = −4.95; p < 0.001) and Total Time (t-test (52) = 7.12; p < 0.001). HC took less time to finish the test (mean = 457.3 ± 73.60) and obtained a better overall score (mean = 12.3 ± 1.07). Furthermore, a substantial difference in the FAB score, a standardized neuropsychological test for EFs, showed that HC achieved a higher performance (17.46 ± 1.003).

	PwPD Mean ± SD	**HC Mean ± SD**	**Group Comparison (*p-Value*)**
EXIT 360° Total Score	10.5 ± 1.58	12.3 ± 1.07	**<0.001**
EXIT 360° Total Time	716.4 ± 174.19	457.3 ± 73.60	**<0.001**
FAB	15.94 ± 2.33	17.46 ± 1.003	**0.006**

Table 3. 5: Comparison of scores at EXIT 360°. SD, standard deviation; PwPD, patients with Parkinson's disease; HC, healthy controls; FAB = Frontal Assessment Battery (in bold, statistically significant value).

Usability: Figure 3.10 displays the average usability score of both groups at the SUS. The comparison of the usability scores between the PwPD and HC revealed no statistically significant difference (t-test (52) = -1.09; p = 0.279). While HC displayed a mean score of 80 ± 11.22, PwPD provided a mean score of 76.94 ± 9.18. In accordance with the scale's score acceptability ranges (cut off = 68) and adjective ratings, both scores show a satisfactory level of usability, included between "good" and "excellent".

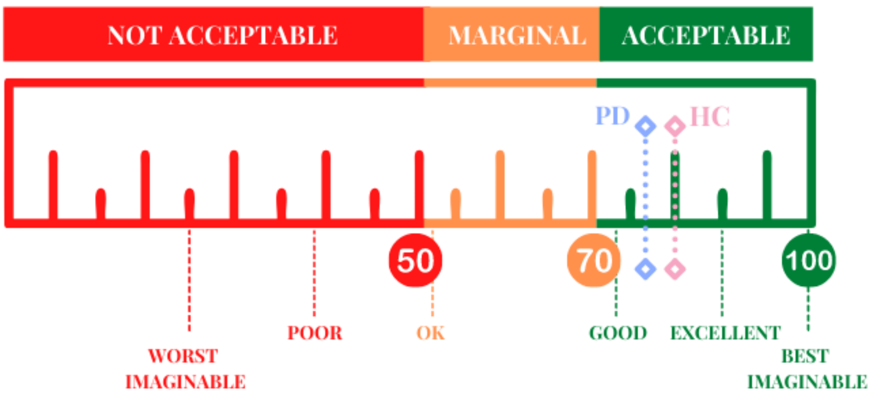


Figure 3.10: A graphic representation of the SUS score. PD, Parkinson's disease; HC, healthy control.

Specifically, more than 74% of PwPD and 92% of HC had scores over the cut-off (cut-off = 68). In addition, 29.6% of PwPD rated EXIT 360° as "OK," 59.3% as "good," 7.4% as "excellent," and 3.7% as "best imaginable" based on the adjective rating (Lewis & Sauro, 2009). Regarding HC, 3.7% of participants assessed EXIT 360° as "OK", followed by ratings of "good" (59.3%), "excellent" (18.5%) and "best imaginable" (18.5%), with only one subject giving a low score. Finally, all participants gave positive and encouraging scores for the two primary factors influencing the user experience. No significant differences appeared between the groups in terms of usability (t-test (52) = -1.96; p = 0.055) or learnability (t-test (52) = 1.89; p = 0.064). In particular, PwPD displayed a mean usability score of 2.98 ± 0.47 (vs 3.24 ± 0.50) and a learnability score of 3.37 ± 0.63 (vs 3.06 ± 0.59). Only 11.1% of patients and 3.7% of controls showed low usability scores (<2.5), and only 3.7% of both groups had low learnability scores.

User Experience: Both PwPD and HC scored well on the first Flow Short Scale item regarding their perceived level of ability when completing EXIT 360° (median = 5, IQR = 4-5), without a significant difference between the groups (U test (52) = 363; p = 0.978). Additionally, the other two items allowed for evaluation of EXIT 360°'s level of challenge according to participants' abilities, as "balance/appropriate" (median = 3; IQR = 3), without significant difference between the two groups (U test (52) = 326; p = 0.191).

Table 3.11 demonstrates that there are no significant differences between the two groups for the subscale "enjoyment" of the IMI, which received high values (≥4) in all items.

	PwPD Median (IQR)	**HC Median (IQR)**	**Group Comparison (*p-Value*)**
Boring	5 (5)	5 (5)	0.107
Enjoyable	4 (4–5)	5 (4–5)	0.113
Activating	5 (5)	5 (5)	0.28
Funny	4 (3.5)	4 (4–4.5)	0.81

Figure 3.11: Comparison of scores at the subscale enjoyment of IMI. IQR, interquartile range; n, number; PwPD, patients with Parkinson's disease; HC, healthy controls.

Specifically, Figure 3.12 shows the percentage relating to all four items of subscale enjoyment of the IMI, comparing two groups. The figure shows that both PwPD and HC considered EXIT 360° activating, funny and enjoyable. No participant evaluated EXIT 360° as boring.

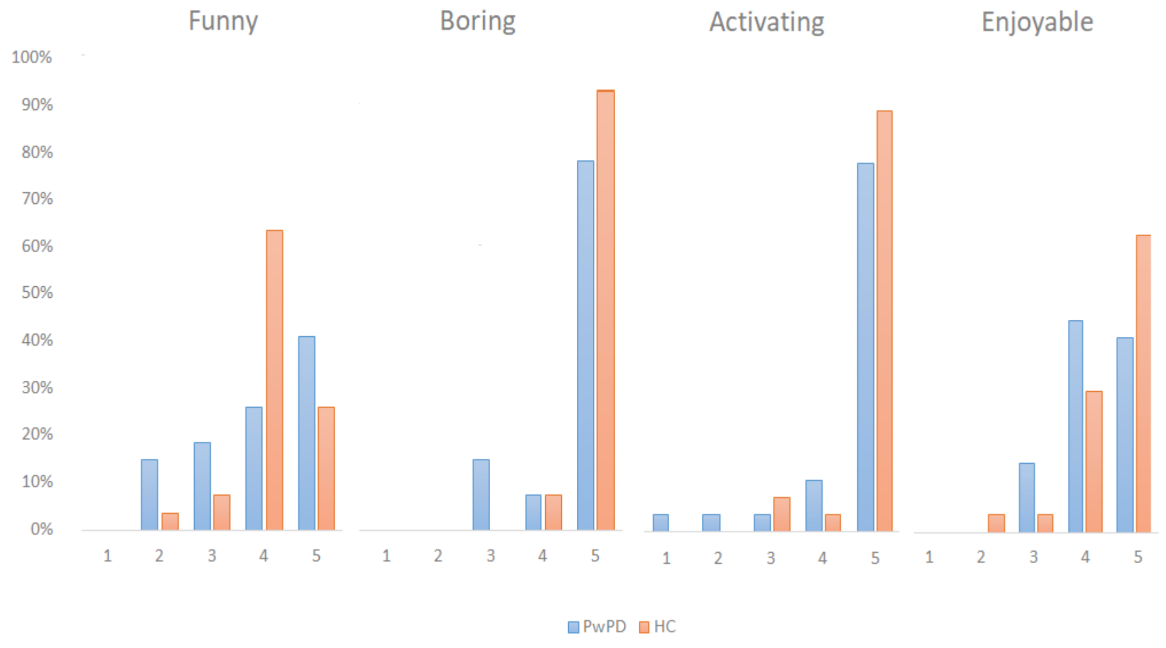


Figure 3.12: Graphic representation of the Intrinsic Motivation Inventory domains, comparing patients and controls

Table 3.6 displays good results in all ICT—SOPI dimensions. A significant difference appeared only in the "engagement" domain (t-test (52) = -3.44; p<0.05). Regarding the "negative effects" domain, few participants (three PwPD and three HC) reported the occurrence of minor negative effects (score < 3), such as vertigo or nausea.

	PwPD Mean ± SD	**HC Mean ± SD**	**Group Comparison (*p*-Value)**
Spatial Presence	3.11 ± 0.83	3.47 ± 0.48	0.054
Engagement	3.43 ± 0.54	3.9 ± 0.47	**0.001**
Ecological Validity	4.29 ± 0.61	4.49 ± 0.37	0.149
Negative Effects	1.29 ± 0.42	1.2 ± 0.26	0.361

Table 3.6: Comparison of scores in ICT—SOPI dimensions. SD, standard deviation; PwPD, patients with Parkinson's disease; HC, healthy controls (in bold, statistically significant value).

Figure 3.13 shows participants' good and promising scores in the "spatial presence" and "ecological validity" domains, divided according to two groups. First, the majority of individuals (70.4 of PwPD and 88.9 of HC) had high ratings for spatial presence (≥3—e.g., ("*I felt I could interact with the environment shown*"). Additionally, 92.6% of patients and all HC agreed that EXIT 360° had good ecological validity ("*I had the feeling that the environment shown was part of the real world*"), with most of the participants (96.3% and 85.2%, respectively) that gave high scores (≥4).

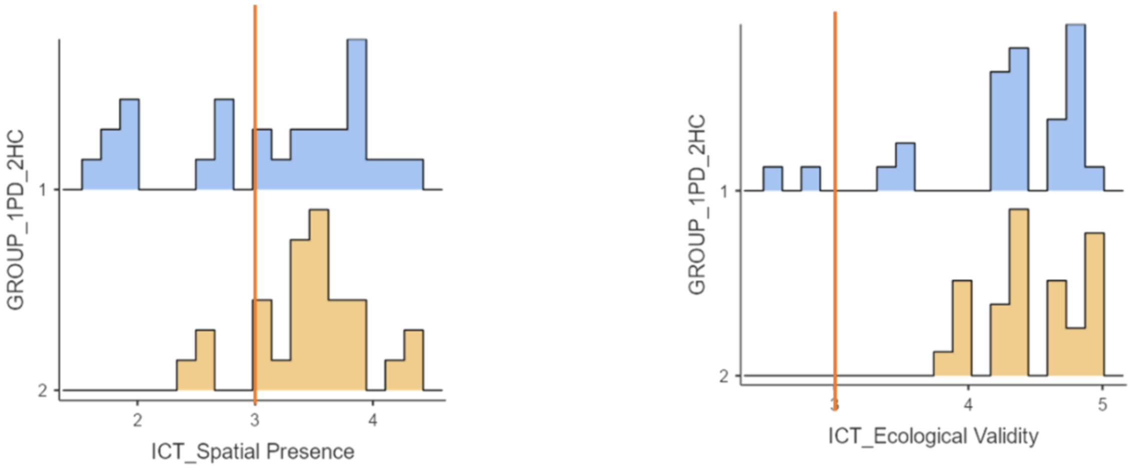


Figure 3.13:Graphic representation of two ICT—SOPI dimensions. The orange lines indicate a neutral score.

Finally, despite the significant difference in the "engagement" domain, the majority of participants demonstrated high levels of engagement while completing EXIT 360° (≥3—e.g., "*I would have liked the experience to continue*", except for six patients and one control.

The UEQ questionnaire showed positive evaluation (>0.8) in all 26 items in both groups. Figure 3.14 shows good scores obtained by PwPD in all UEQ scales according to the questionnaire's range (range between −3, horribly bad, and +3, extremely good).

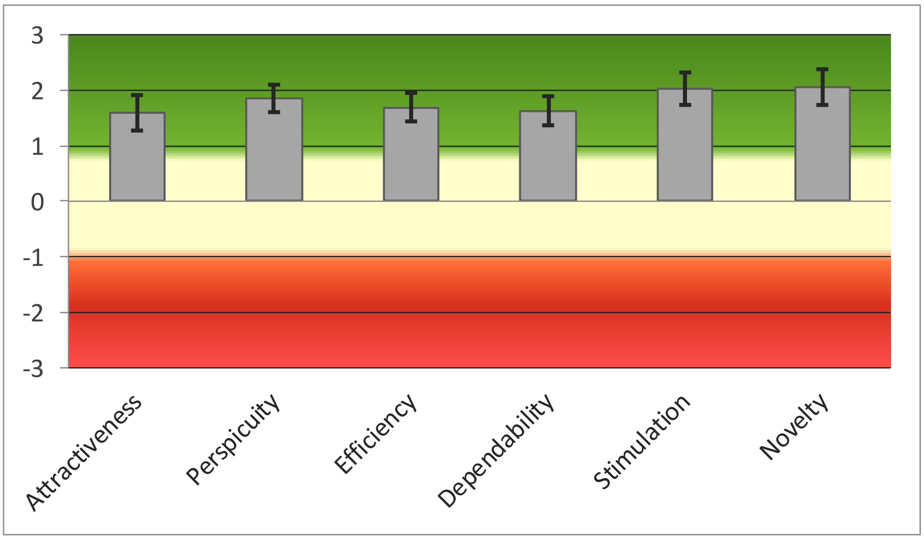


Figure 3.14: Graphic representation of scores of the six UEQ scales.

Table 3.7 details the high mean scores of all scales (regarding PwPD) with their corresponding good internal consistency values (Alpha-coefficient > 0.7) (Cronbach, 1951).

	Mean	**SD**	**Confidence Interval**		**Alpha-Coefficient**
Attractiveness	1.593	0.846	1.273	1.912	0.89
Perspicuity	1.852	0.655	1.605	2.099	0.81
Efficiency	1.694	0.681	1.438	1.951	0.72
Dependability	1.630	0.695	1.368	1.892	0.78
Stimulation	2.028	0.776	1.735	2.321	0.79
Novelty	2.056	0.853	1.734	2.377	0.93

Table 3.7: Scores of the six UEQ scales. SD, standard deviation.

Table 3.8 shows good scores in all UEQ scales as well as the pragmatic and hedonic quality dimensions of PwPD and HC, including the comparison between the two groups. In detail, the scales of the UEQ can be divided into two categories: hedonic quality (non-task-related quality aspects—originality and stimulation) and pragmatic quality (perspicuity, efficiency, dependability).

	PwPD Mean (SD)	**HC Mean (SD)**	**Group Comparison (*p-Value*)**
Attractiveness	1.59 (0.85)	1.81 (1.13)	0.430
Perspicuity	1.85 (0.66)	2.01 (0.75)	0.416
Efficiency	1.69 (0.68)	1.73 (0.84)	0.859
Dependability	1.63 (0.70)	2.14 (0.86)	**0.020**
Stimulation	2.03 (0.78)	2.08 (0.98)	0.818
Novelty	2.06 (0.85)	2.46 (0.68)	0.058

Pragmatic Quality	1.73 (0.59)	1.96 (0.74)	0.204
Hedonic Quality	2.04 (0.72)	2.27 (0.82)	0.275

Table 3.8: Comparison of scores in UEQ scales and dimensions. SD, standard deviation; PwPD, patients with Parkinson's disease; HC, healthy controls (in bold, statistically significant value).

As in Study 1, the means of each UEQ scale of PwPD were then compared to pre-existing values from a benchmark dataset (Schrepp et al., 2017). Results revealed that the scale's novelty and stimulation obtained an excellent rate, falling within the range of the 10% best results (Figure 3.15). Additionally, the Attractiveness, Perspicuity, Efficacy, and Dependability scales received a good evaluation: "10% of results better, 75% of results worse."

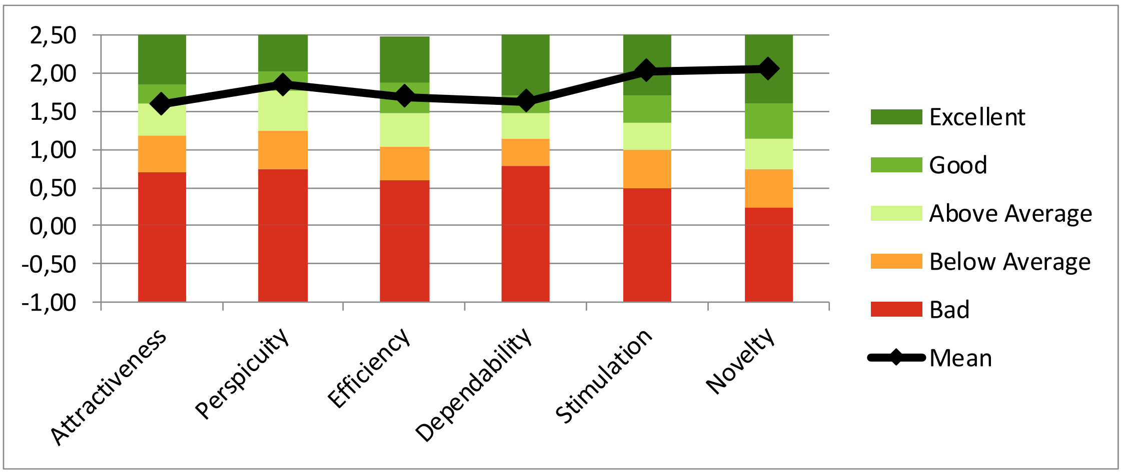


Figure 3.15: Comparison between means of each UEQ scale and values from a benchmark dataset.

Correlation: Pearson's correlation revealed no significant linear relationship between the SUS total score and education (r = 0.078; p = 0.576). However, there was a correlation between age and the SUS total score (r =-0.401; p < 0.05). Additionally, Pearson's correlation has demonstrated that there is no significant relationship between the SUS total score and the technological expertise measured by the ad hoc competence questionnaire, either for patients (r = 0.340; p = 0.082) or controls (r = 0.244; p = 0.221). PwPD did not demonstrate a linear correlation between the SUS total score and the three ICT—SOPI domains of spatial presence (r = 0.293; p = 0.138), engagement (r = 0.361; p = 0.064), and ecological validity (r = 0.282; p = 0.154). Similar results were achieved by HC, except for ecological validity, which showed a strong correlation with the usability score (r = 0.422; p = 0.028). Additionally, data revealed a significant negative association (r = -0.325; p 0.05) between the SUS total score and the ICT—SOPI "negative effects" effect only in PwPD. Finally, statistical analysis demonstrated no correlation between usability score and the three main dimensions of UEQ, respectively attractiveness (r = 0.168; p = 0.224), pragmatic quality (r = 0.196; p = 0.157), and hedonic quality (r = 0.250; p = 0.069).

Conclusion

Study 2 has shown encouraging and valuable results on the usability and quality of user experience of EXIT 360°, reflecting those of Study 1.

In line with the previous study, PwPD successfully completed EXIT 360° without experiencing any noteworthy adverse effects. PwPD provided a positive overall impression of EXIT 360° describing it as usable, easy to learn, clear, pleasant, engaging, funny and not boring. As widely demonstrated in Study 1, PwPD supported that EXIT 360° can be considered an innovative, efficient, straightforward, and fast technological solution, with great hedonic quality regarding stimulation (challenging, exciting and interesting) and novelty. The good spatial presence, excellent ecological validity and irrelevant adverse effects have been confirmed as key features of EXIT 360°. Interestingly, the two groups obtained similar scores in most of the analyzed dimensions, except for "dependability" and "engagement", in which PwPD provided a lower score. The dependability domain was influenced by the item "meets expectations" because patients reported having negative expectations due to traditional long and complex evaluations and claimed to be pleasantly surprised. As regards "engagement", only six patients showed low scores. However, they claimed that "the evaluation had a correct duration but that they would have had no problems continuing". Finally, technological expertise or education level did not influence the encouraging results. Only the age variable showed a negative correlation with the usability score; however, older people (both patients and controls) were able to complete the evaluation with some instructions.

Usability Studies: Conclusion, Limitations and Future Perspectives

In this Chapter, we have focused on promising and interesting results related to the usability and user experience quality of EXIT 360°, in line with the widely demonstrated critical role of these two components in the development, validation and use of VR-based tools as EXIT 360°.

The studies have first shown that both clinical and healthy populations successfully carried out EXIT 360°. At the baseline, the sample showed low technological expertise regarding perceived familiarity and technology competence. For example, study 2 showed that only 7.4% of PwPD and 18.5% of HC claimed good technological competence. Despite the low familiarity and competence with the technologies, all subjects were able to complete the test. As a result, EXIT 360° appeared as a viable tool that clinicians could use even with patients without prior exposure to technology. Additionally, both studies have converged in supporting that EXIT 360° has a good to excellent overall usability. Specifically, EXIT 360° has been seen by both HC and PwPD as a usable and easy-to-learn technological tool suitable for their skill levels (Borgnis, Baglio, Pedroli, Rossetto, Meloni, et al., 2022).

Due to this interesting usability finding, we were able to draw the following conclusions about EXIT 360°: it demonstrated high effectiveness (i.e., good ability for users to achieve goals), efficiency (i.e., few users' efforts to achieve the goal), and satisfaction (i.e., good users' thoughts about their interaction with the system) (Iso, 1998). Therefore, it is conceivable to argue that any subject's low performance will not depend on technological problems. These optimistic results were unaffected by technological expertise or education level. As demonstrated, only the age variable negatively correlated with the usability score. However, older adults (both patients and controls) were able to finish the assessment with some suggestions. Overall, these findings suggest that our system wouldn't need to be modified. In contrast to the only prior usability study on PD and VR-based instruments (Pedroli et al., 2013), these results were encouraging. Data from that study revealed that healthy participants provided good usability for the VR-based tool, whereas PwPD participants demonstrated that the instrument requires more than just an improvement to be considered usable.

Promising results were also obtained by evaluating the user experience quality of EXIT 360° in both groups. All participants supported a positive general impression of the EXIT 360°, considering it attractive (e.g., pleasing, friendly, and enjoyable), activating, funny and not dull. Furthermore, EXIT 360° demonstrated good pragmatic qualities by being: (1) quick, efficient, practical, and organized (efficiency); (2) perceptive, understandable, easy to learn, and straightforward (perspicuity); and (3) predictable, supportive, and secure (dependability). Except for dependability, where patients received a lower score than controls, the two groups shared scores for all these factors. However, the item "meets expectations" had an impact on this domain since patients reported they had negative expectations due to customarily drawn-out and complicated evaluations but were pleasantly surprised. Also, EXIT 360° revealed excellent hedonic quality in terms of stimulation (valuable, exciting, engaging, and motivating) and novelty (creative, innovative, inventive). Interestingly, results showed that "attractiveness", "dependability", "stimulation", and "novelty" scales obtained excellent evaluation compared to existing values from benchmark data (Schrepp et al., 2017). Finally, everyone who took part said that EXIT 360° was an engaging and challenging tool with good spatial presence ("*I felt I could interact with the environment shown*") and excellent ecological validity ("*I had the feeling that the environment shown was part of the real world*"). In terms of the "engagement" domain, most participants demonstrated good levels of engagement while completing EXIT 360° (e.g., "*I would have liked the experience to continue*"); the only six patients who provided low scores did so in support of the statement that "*the evaluation had a correct duration but that they would have had no problems continuing*". The high levels of engagement, enjoyment, and attractiveness of EXIT 360° could increase users' motivation and participation while reducing the fear and anxiety typically associated with neuropsychological testing. Lastly, only a small number of patients demonstrated the existence of irrelevant side effects that disappeared quickly, such as nausea or dizziness. It is a significant result since the feeling of cybersickness can lead to unpleasant experiences for users, impacting their performance and significantly decreasing the test results' validity.

Overall, the two studies have contributed important findings to the development and validation of EXIT 360° in line with previous research that demonstrates the importance of usability and user experience in creating digital content (Aubin et al., 2018; Schultheis & Rizzo, 2001). In other words, this Chapter has shown that EXIT 360° can satisfy all of the following significant criteria: good usability, absence of side effects, high sense of presence, ecological validity, enjoyment, and attractiveness.

Limitation and Future perspective: The present work has some limitations. Firstly, the technology employed was entry-level: the 360° mobile-powered devices currently available on the market (e.g., Oculus Quest) have far greater quality, allowing for better 360° image quality and a better, more realistic experience. Furthermore, to date, EXIT 360° is still a preliminary prototype that requires additional validation steps to be a reliable and standardized instrument for evaluating EFs. For this reason, it will be necessary to assess the convergent validity of EXIT 360°, comparing it with the standard neuropsychological battery for executive functioning and its effectiveness in discriminating between HC and PwPD (i.e., how well it can distinguish between PD and HC). The evaluation of these crucial psychometric properties of EXIT 360° will be the main topic of the following chapters.

4

4. EXIT 360° - Convergent Validity Study

The usability assessment represented the first step in validating EXIT 360° as an instrument tool for executive functioning. The excellent results of usability studies allow us to carry on the validation process of EXIT 360°, focusing on the other two essential psychometric proprieties: convergent and construct validity (Krabbe, 2017). This Chapter will focus on the study's main results to assess convergent validity.

The concept of "convergent validity" means how closely the new scale or tool is related to other variables and measures of the same construct (Krabbe, 2017). In other words, it assumes that tests based on the same or similar constructs should be highly correlated. In this context, one of the most used methods is to correlate the scores between the new assessment tool and others that are claimed to measure the same construct (Chin & Yao, 2014). For this purpose, we have compared EXIT 360° with standardized traditional neuropsychological tests for executive functioning.

As widely said in previous chapters, EXIT 360° was conceived and designed as a novel task for EFs that requires participants to complete routine tasks in 360° real-life settings. In particular, EXIT 360° was born to provide an evaluation of multiple EFs, allowing one to obtain a real executive status of subjects. A multicomponent evaluation appeared us the best solution since the term "EFs" involves a wide range of neurocognitive processes and behavioral skills, such as planning, attention, control inhibitor, cognitive flexibility, and working memory (Chan et al., 2008). In detail, the EXIT 360° development project aimed at creating an innovative tool able to evaluate the following EFs:

Selective attention is the cognitive process that allows for focusing on a particular object in the environment for a certain period. It enables one to focus attention on relevant stimuli while ignoring unimportant stimuli in the environment (McLeod, 2018). It is a crucial process since there is a limit to how much information can be processed at a given time.

Divided Attention is the ability to perform more than one activity at a time. "Having a conversation while walking" or "listening to music while shopping" are examples of divided attention.

Cognitive flexibility is the ability to change behavioral patterns based on feedback received. In other words, it allows for revising plans due to obstacles, setbacks, new information, or mistakes.

Set shifting represents the ability to switch freely and fast from one situation, activity, or aspect of a problem to another in reaction to internal or external cues. Patients with dorsolateral prefrontal injury showed impairments in set-shifting, appearing unable to modify a response strategy according to the requests of a task. Therefore, these patients could make many perseverations (Milner, 1963)

Working memory is a cognitive system with limited capacity that can contain information temporarily (Miyake & Shah, 1997). It allows for holding data in memory in a heightened state of availability for use in performing many complex cognitive tasks (Cowan et al., 2014). Working memory differs from short-term memory because it permits manipulation of information that has been stored there, whereas short-term memory only refers to the temporary storage of information.

Inhibition Control consists of the ability to think before acting, to hold back on saying or doing something to allow the time to evaluate a situation and the potential impact of what is said or done. Additionally, it allows for inhibiting previously learned responses and controlling the interference effect of distracting stimuli, which are irrelevant to continue applying the plan. Patients with anterior girdle injuries generally exhibit errors of this nature in their cognitive function (Bush et al., 2000).

Planning is the ability to establish and coordinate the actions and components required to carry out an intention and reach a goal (Lezak et al., 2004). Planning tasks require the subject to predict the goal to be achieved, break down the action into intermediate steps, order the steps and monitor the task's execution in relation to the predetermined outcome. Patients with dorsolateral prefrontal injury showed difficulties in this cognitive ability (Shallice, 1982).

Reasoning is the ability to deliberately use logic by making inferences from new or existing data.

Problem-solving is the process of examining the given information to identify every possible resolution. It involves defining a problem, determining the cause, locating, and choosing potential solutions, as well as putting those solutions into action.

Decision-making is the ability to select an action from several alternatives following the individual's goals and motivations.

For evaluating convergent validity, we have chosen an assessment protocol that included standard neuropsychological paper-and-pencil tests that have historically been used to assess these EFs.

Material and Methods

Participants: 77 healthy volunteers were recruited at IRCCS Fondazione Don Carlo Gnocchi in Milan, based on the inclusion criteria: (a) age ranged 18-90 years old; (b) education level ≥5; (c) no cognitive impairment as measured MoCA test; and (d) ability to provide written, signed informed consent. Exclusion criteria were: (a) overt visual or hearing impairment or visual hallucinations or vertigo and (b) systemic, psychiatric, or neurological diseases.

Procedures: Participants underwent a one-session evaluation that involved three of four phases described in the section "EXIT 360° Validation", in line with the study's aim. Briefly, the evaluation consisted of 1) a traditional neuropsychological assessment (Pre-Task Evaluation), 2) EXIT 360° Session, and 3) a brief Post-Task Evaluation (only usability assessment). The traditional neuropsychological battery involved, beyond the MoCA test for global cognitive functioning, the main standardized paper-and-pencil tests for the evaluation of executive functioning:

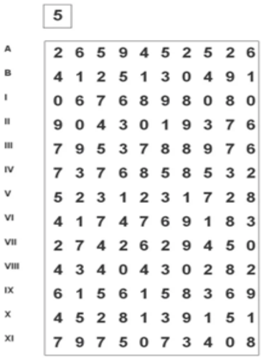
(1) Attentive Matrices (Spinnler & Tognoni, 1987a)
(2) Trail Making Test (Reitan, 1958, 1992a)
(3) Progressive Matrices of Raven (Caffarra et al., 2003; Raven, 1947)
(4) Frontal Assessment Battery (FAB) (Appollonio et al., 2005; Dubois et al., 2000)
(5) Stroop Test (Golden & Freshwater, 1978; Stroop, 1935; Venturini et al., 1983)
(6) Phonemic verbal fluency task (F.A.S.) (Novelli et al., 1986)
(7) Digit Span Backward (Monaco et al., 2013).

For greater clarity, we will describe the purpose (i.e., which EFs are evaluated), structure (i.e., how it is administered), and scoring (i.e., how the scores are calculated) for each test. The raw scores obtained by the test subject were corrected for gender, age and education according to available normative Italian and subsequently converted to the corresponding Equivalent Score, graded on a 5-level scale: 0 = Deficit, 1 = Borderline, ≥2 = normality.

Attentional matrices (Spinnler & Tognoni, 1987a):

<u>Purpose</u>: Selective visual attention:

<u>Structure</u>: 3 matrices containing numbers from 0 to 9 (13 lines – 10 numbers per line) arranged randomly. Depending on the matrix, the subject must cross specific numbers ("5" - "2-6" - "1-4-9") in 45 seconds. The first two lines (A and B in Figure) are "tests" to ensure the patient understands the task and the subject must complete them before starting timing.

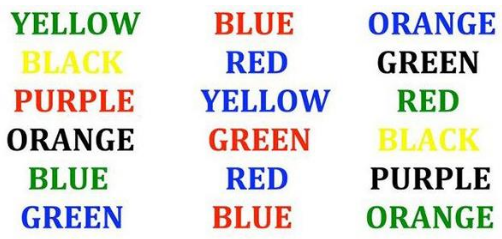


<u>Scoring:</u> The total score is calculated by the sum of the targets correctly identified within the three matrices (excluding those of the test lines).

Stroop test (Golden & Freshwater, 1978; Stroop, 1935; Venturini et al., 1983):

<u>Purpose</u>: visual attention and inhibitory control.

<u>Structure</u>: The test consists of three consecutive parts that the subject must perform as quickly and accurately as possible: 1) read a list of color names on a sheet; 2) say the names of the colors of some colored spots on the sheet; 3) name the color of the ink with which the names of the colors were written (see Figure). For example, if the words "green" are written in yellow, the subject must pronounce "yellow". In this case, the subject must inhibit the automatic tendency to read the words, paying attention to the ink.



<u>Scoring:</u> The examiner must collect subjects' errors (accuracy) and the time taken to complete (processing speed) each part. Then, he must calculate two global scores by summing the errors and times of the test. Finally, two standardized formulas will be used to calculate the two final raw scores.

Verbal Phonological Fluency (Novelli et al., 1986)

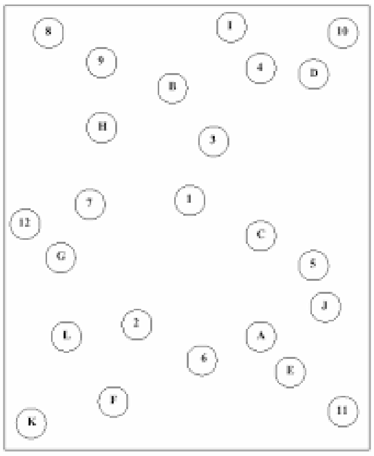
<u>Purpose</u>: ability to access the lexicon and lexical organization.

<u>Structure</u>: the subject must say all the words that begin with a certain letter in a minute. The test includes three letters (for example, in the version of Novelli: F-A-S). Performance is typically impaired in the case of left frontal injuries.

<u>Scoring:</u> The total score is calculated by the sum of the targets correctly identified within the three matrices.

Trail Making Test (TMT) (Reitan, 1958, 1992b, 1992a):

Purpose: visual attention, visual-spatial exploration, and cognitive flexibility.

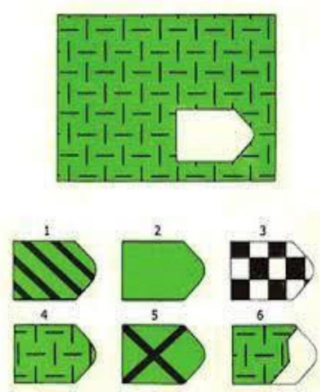


Structure: The test consists of connecting 25 consecutive stimuli, circled, and printed on a sheet of A4 paper in the shortest possible time. The test consists of two parts: Part A: join a series of numbers in ascending order; Part B: connect all the stimuli on the sheet (numbers and letters), alternating a number and a letter following the increasing order of the numbers and the alphabetical order of the letters (1-A-2-B)

Scoring: The scoring of Part A (i.e., time to complete Part A) and Part B provides an assessment of visual attention. In addition, Form B requires the patient to switch cognitive sets between numbers and letters, allowing to evaluate the ability of task switching (an index of cognitive flexibility). Additionally, the scoring system includes a score "B-A", calculated as the difference between the two times and used to assess the ability to transition between tasks independently of any visual attention deficits.

Progressive Matrices of Raven (Caffarra et al., 2003; Raven, 1947).

Purpose: logical-abstract reasoning and spatial skills, regardless of previously learned notions

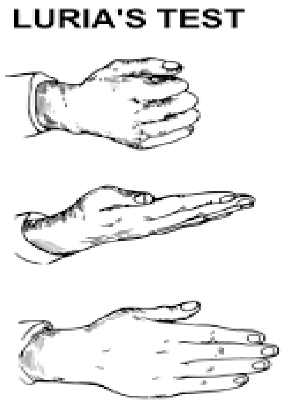


Structure: The test consists of four series (A, B, C, D) of 12 figures, each of increasing complexity. Each figure is missing a piece, and the subject will have to choose from a series of drawings, the one that allows him to complete the figure. The Progressive Colored Matrices are used for children under the age of 8, adults of low intellectual level or the elderly; they include series A, B, and an intermediate Ab series between them.

Scoring: The total score is calculated by the sum of the figure correctly identified within the three or four series.

Frontal Assessment Battery (FAB) (Appollonio et al., 2005; Dubois et al., 2000):

Purpose: a screening battery including a series of cognitive and behavioural tests that allow for evaluating different EFs.



Structure: The subject must perform six tasks: 1) Similarities: “in what are they similar…? (Conceptualization - abstraction) 2) Lexical fluency in phonemic modality (cognitive flexibility) 3) Luria's “punch-cut-flat” motor series (programming, planning and organization of behavior – see Figure); 4) Response to conflicting instructions: "*tap twice when I tap once*" (sensitivity to interference) 5) Go-No-Go task: "*don't tap when I tap twice*" (inhibitory control, impulsiveness) 6) Prehension Behavior (environmental autonomy). These tests allow us to hypothesize deficits in different areas of the frontal, prefrontal (dorsolateral or orbitofrontal) and anterior cingulate cortex.

Scoring: Each task can be evaluated from 0 to 3, and the sum of the tasks scores calculates the total score.

Digit Span Backward (Monaco et al., 2013)

Purpose: working memory

Structure: The examiner reads a numerical sequence (one number per second), and then the subject must repeat the sequence in reverse (i.e., from the last number heard to the first). For example, the sequence “6-2-9” must become “9-2-6”. If the subject correctly repeats the series, the examiner continues with the following sequence, which is one number longer than the previous one. On the contrary, the examiner will read another sequence containing the same amount of numbers. The test continues until the subject fails two identical sequences.

Time 1	Time 2
6-2-9	4-1-5
3-2-7-9	4-9-6-8
1-5-2-8-6	6-1-8-4-3
5-3-9-4-1-8	7-2-4-8-5-6
8-1-2-9-3-6-5	4-7-3-9-1-2-8
9-4-3-7-6-2-5-8	7-2-8-1-9-6-5-3

Scoring: The score is given by the amount of numbers of the last correctly repeated sequence.

Statistical Analysis: Descriptive statistics involved mean and standard deviation (SD) for continuous measures and frequencies, percentages, median, and interquartile range (IQR) for categorical variables. The normality of data distribution was assessed using the Kolmogorov-Smirnov test. Pearson’s correlation (or Spearman correlation according to variables) was conducted to verify a possible relationship between the scores of neuropsychological tests and EXIT 360° (Total Score, Total Time, and subtasks scores). Additionally, total scores of EXIT 360° and usability score were compared using Pearson's correlation to confirm that there was no effect of technological usability on performance. Additionally, we evaluated the association (with univariate and multiple linear regression) between EXIT 360° variables and demographic characteristics to verify the possible influence of socio-demographic features on the results of EXIT 360°. Jamovi 1.6.7 software was used to conduct all statistical analyses. A statistical threshold of $p<0.05$ was considered statistically significant.

Results

Participants: Table 4.1 reports the demographic and clinical characteristics of the whole sample. The subjects (n=77) are mostly female (M:F= 29:48) with a mean age of 53.2 years (SD=20.40, range=24-89) and age of education nearly 13 (IQR=13-18, range 5-18). No cognitive impairment appeared in the sample (MoCA correct score=25.9±2.62).

		Subjects [N=77]
Age (years)	*Mean (SD)*	53.2 (20.40)
Sex (M:F)		29:48
Age of education (years)	*Median (IQR)*	13 (13-18)
MoCA_raw score	*Mean (SD)*	26.9 (2.37)
MoCA_correct score	*Mean (SD)*	25.9 (2.62)

Table 4. 1: Demographic and clinical characteristics of the whole sample. M=Male; F=Female; SD: Standard Deviation; IQR= Interquartile Range; N=number; MoCA= Montreal Cognitive Assessment.

Traditional Neuropsychological Assessment: The average scores (raw and corrected scores) of the neuropsychological tests are shown in Table 4.2, together with the corresponding cut-off for normality (equivalent score≥2). Specifically, all study participants obtained scores within the normal range on all conventional neuropsychological tests for executive functions.

Neuropsychological tests	**Raw Score** *Mean (SD)*	**Corrected Score** *Mean (SD)*	**Cut-off of normality**
Trail Making Test – Part A	37.2 (22.9)	35.1 (19.3)	≤68
Trail Making Test – Part B	94.1 (58.9)	90.6 (48.4)	≤177

Trail Making Test – Part B-A	56.9 (42.3)	57 (34.2)	≤111
Verbal Fluency Task	41.6 (11.1)	37.9 (9.46)	≥23
Stroop Test – Errors	0.68 (1.09)	0.62 (1.13)	≤2.82
Stroop Test – Time	22.6 (15.5)	23.8 (11.5)	≤31.65
Frontal Assessment Battery	17.6 (1)	17.7 (0.85)	≥14.40
Digit Span Backward	4.77 (0.99)	4.56 (0.97)	≥3.29
Attentive Matrices	54.3 (5.53)	48.6 (6.43)	≥37
Progressive Matrices of Raven	32.3 (3.63)	32.3 (3.2)	≥23.5

Table 4. 2: Scores of Neuropsychological Assessment. SD: Standard Deviation

EXIT 360°: All study participants have completed the seven subtasks, obtaining only one point for wrong answers or two points for correct ones. The subjects' scores (%) on each subtask are shown in Figure 4.1.

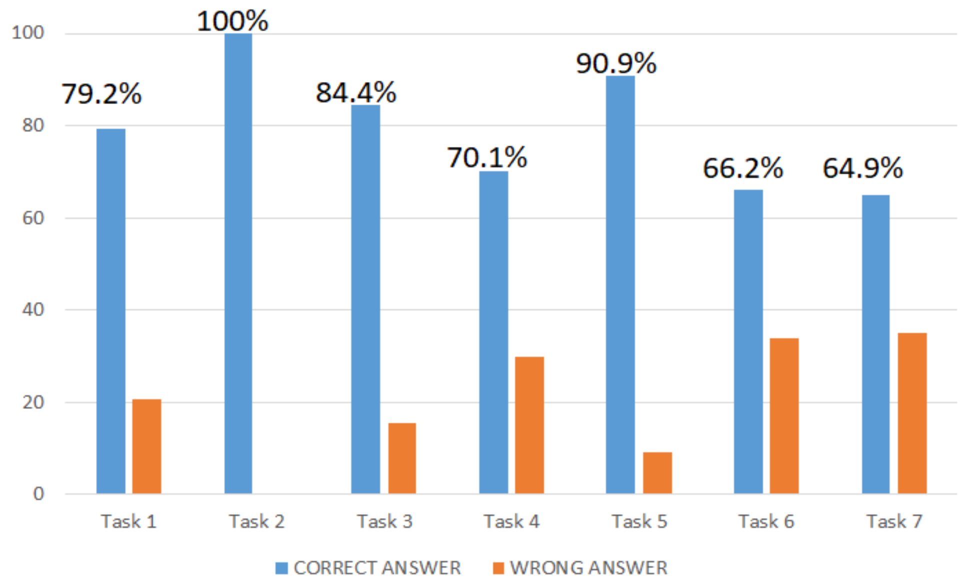


Figure 4.1: scores (%) on seven subtasks

Overall, the descriptive analysis revealed that healthy volunteers received, on average, a EXIT 360° Total Score of 12.6 (±1.02; range= 10-14), with 88.3% of subjects that got a score ≥ of 12. Additionally, participants took about 8 minutes (Total Time mean=480 seconds ±130 sec; range = 192-963 sec) to complete the whole task.

EXIT 360° Total Score and Demographic characteristics: The univariate linear regression showed a significant impact of age (β=-.451, $p<.001$; R^2=.203) and education ($p<.001$; R^2=.300) on EXIT 360° Total Score, but not of gender (β=-.0980; p=.680; R^2=.002). Specifically, regarding education, there was a significant difference between participants with low levels of education (5 years) and medium-high ones, respectively, 13 (β=1.635, $p<.001$), 16 (β=1.962, $p<.001$) and 18 (β=1.923, $p<.001$). Furthermore, Pearson's correlation revealed a significant negative correlation between age and EXIT 360° Total Score (r=-.451; $p<.001$). Regarding EXIT 360° Total Time, univariate linear regressions showed no significant impact of all demographic characteristics on the time variable ($p>.05$). The multiple linear regression (R^2=.342) confirmed the effect of education on EXIT 360° Total Score ($p<.05$) but not the impact of age that showed only a tendency to significance (β=-.239, p=.051). Finally, the variable "sex" did not impact the EXIT 360° Total Score (β=-.127, p=.528).

Correlation between Neuropsychological Tests and EXIT 360°: Table 4.3 shows correlations (Pearson's correlation) between traditional paper and pencil neuropsychological tests and the two scores of EXIT 360°.

	EXIT 360° Total Score	**EXIT 360° Total Time**
Montreal Cognitive Assessment	**0.48****	**-0.31***
Progressive Matrices of Raven	**0.44****	-0.11
Attentive Matrices	**0.26***	**-0.23***
Frontal Assessment Battery	**0.41****	-0.04
Verbal Fluency Task	**0.54****	-0.13
Digit Span Backward	**0.32***	**-0.23***
Trail Making Test – Part A	**-0.45****	0.14
Trail Making Test – Part B	**-0.34***	**0.27***
Trail Making Test – Part B-A	**-0.23***	**0.29***
Stroop Test – Errors	**-0.32***	**0.25***
Stroop Test – Time	**-0.45****	**0.28***

Table 4.3: Correlation between EXIT 360° scores and Neuropsychological Assessment. In bold, statistically significant scores. *p<.05; **p<.001

In detail, a significant correlation appeared between the EXIT 360° Total Score and all neuropsychological tests. Additionally, data showed a relationship between EXIT 360° Total Time and various assessment tests, particularly timed ones (e.g., Trail Making Test, Stroop Test, and Attentive Matrices). Moreover, data showed no relationship between EXIT 360° Total Score and EXIT 360° Time (p=0.587).

Finally, Table 4.4 reveals the significant correlation between traditional neuropsychological tests and the score (Spearman's correlation) and reaction time (Pearson's correlation) of seven subtasks.

	Task 1		**Task 2**		**Task 3**		**Task 4**		**Task 5**		**Task 6**		**Task 7**	
	S	**T**	**S**	**T**	**S**	**T**	**S**	**T**	**S**	**T**	**S**	**T**	**S**	**T**
PMR									x		x		x	
AM						(x)							x	x
FAB							x				x		x	
V.F.T.											x		x	
DS		x							x		x	x	x	
TMT – A									x		x		x	x
TMT – B		x							x		x			
TMT B-A		x							x					
ST_E									x	x				

Table 4.4: Correlation between subtask scores and Neuropsychological Assessment. S=score; T=Time; x =statistically significant scores; (x) = scores tendency to statistical significance

Specifically, data showed significant correlations between:

- Progressive Matrices of Raven and scores of Tasks 5 (r=.241; p<.05), 6 (r=.484; p<0.001) and 7 (r=.296; p<0.05).
- Attentive Matrices and both scores of Task 7 (r=.284; p<.05 – r=-.226; p<.05). Moreover, data showed a tendency to statistical significance for the correlation between this neuropsychological test and Task 3 (r=-.218; p=.057)

- Frontal Assessment Battery and scores of Tasks 4 (r= .254; p<.05), 6 (r=.266; p<.05) and 7 (p=.283; p <.05)
- Verbal Fluency Test and scores of Tasks 6 (r=.489; p<.001) and 7 (r=.438; p<.001)
- Digit Span Backward and scores of Tasks 5 (.251; p<.05), 6 (.341; p<.01) and 7 (r=.303; p<.01). Moreover, results showed a correlation between this test and reaction time of tasks 1 (r=-.269; p<.05) and 6 (r=-.253; p<.05)
- Trail Making Test (TMT) – part A and scores of Tasks 5 (r=-.301; p<.01), 6 (r=-.462; p<.001) and 7 (r=-.299; p<.01). Moreover, results showed a correlation between this test and reaction time of tasks 7 (r=.244; p<.05).
- TMT – part B and respectively the scores of Tasks 5 (r=-.31; p<.01) and 6 (r=-.36; p<.01) and the reaction time of Tasks 1 (r=.333; p<.01)
- TMT – part B-A and respectively the score of Tasks 5 (r=-.259; p<.05) and the reaction time of Task 1 (r=.366; p<0.01)
- Stroop Test – error and both scores of Task 5 (r=-.29; p<.05 – r=.28; p<.05)
- Stroop Test – time and respectively the scores of Tasks 4 (r=-.282; p<.05), 5 (r=-.297; p<.01), 6 (r=-.344; p<.01) and 7 (r=-.329; p<.01) and reaction time of Tasks 1 (r=.339; p<.01) and 7 (r=.286; p<.05)

Usability: The mean value of the usability, measured using SUS, was 75.4±13.2, suggesting an acceptable level of usability according to the scale's score (cut-off=68). Specifically, more than 70% of participants had scores higher than the cut-off. In addition, according to the adjective rating, 35.5% of subjects evaluated EXIT 360° as "Good", 32.9% as "Excellent", and 27.6% as "Best imaginable" (Lewis & Sauro, 2009).

Pearson's or Spearman's correlation showed no significant linear correlation between the SUS total score and the demographic characteristics, respectively, age (r=-0.045, p=0.699) and education (r=-0.096; p=0.405). Moreover, data indicated an absence of significant correlation between the SUS total score and the Total Score of EXIT 360° (r=0.126; p=0.276).

Discussion and Conclusion

Recently, there has been a growing interest in using VR-based solutions for an ecologically valid assessment of executive functioning in numerous clinical populations (Realdon et al., 2019; Serino et al., 2017), able to detect impairments invisible to traditional measurements (Cipresso, la Paglia, et al., 2013; Serino et al., 2017). In this framework, we have used 360° technologies to create EXIT 360°, an innovative diagnostic tool that aims to identify quickly multiple executive deficits, involving participants in a "game for health" where they perform everyday subtasks in 360° domestic environments (Borgnis, Baglio, Pedroli, Rossetto, Meloni, et al., 2021; Borgnis, Baglio, Pedroli, Rossetto, Riva, et al., 2021). After widely describing the promising results of EXIT 360° in terms of usability, this Chapter has focused on demonstrating the good convergent validity of EXIT 360°, comparing it with traditional standardized neuropsychological tests for executive functioning. Indeed, it is well known that a strong positive correlation between a new tool and other instruments designed on the same construct is evidence of the high convergent validity of the new test (Chin & Yao, 2014). Moreover, we have detailed the main characteristics of EXIT 360°, focusing on each subtask and also observing the potential impact of demographic characteristics on indexes.

In our study, we involved a heterogeneous sample of 77 healthy control volunteers that included a wide range of age (24 to 89 years old, that is, from young to elderly adults) and education (5 to 18, that is, from primary school to university). According to the inclusion criteria, participants did not show cognitive or executive impairments, obtaining scores within the normal range on all standard paper-and-pencil neuropsychological tests. As in previous studies, all participants were able to complete the entire task without the psychologist's help. During EXIT 360° session, participants had to complete all subtasks, obtaining only one point for incorrect responses or two points for right ones. The achieved EXIT 360° Total Scores were compared with traditional neuropsychological tests for EFs to evaluate the convergent validity. Data revealed a significant correlation between the EXIT 360° Total Score and all neuropsychological tests for executive functioning. An interesting and promising association was also found between EXIT 360° Total Time and timed neuropsychological tests, like the Trail Making Test, Stroop Test, and Attentive Matrices.

As previously said, a high correlation between indexes of a new test (EXIT 360°) and scores of other standardized instruments that evaluate the same construct (i.e., executive functioning) supported the high level of convergent validity of the new tool. We can conclude that EXIT 360° showed a good convergent validity. Therefore, EXIT 360° can be considered an innovative solution to evaluate several components of executive functioning. In other words, assessing simultaneously selective and divided attention, cognitive flexibility, set-shifting, working memory, reasoning, inhibition, and planning is possible with this tool.

After demonstrating the EXIT 360° convergent validity, we have paid more attention to the main characteristics of the whole task and subtasks. Analysis showed that an evaluation with

EXIT 360° did not require a long administration time; indeed, participants took, on average, about 8 minutes to complete the whole task. It is very important since EXIT 360° can allow an evaluation of multiple components of executive functionality in a short time, overcoming too long neuropsychological evaluation. As regards the accuracy score, that is the EXIT Total Score, most of the participants (over 88%) achieved high scores (≥12). It will be interesting to see if a score of 12 could be a good cut-off value able to differentiate between healthy and pathological groups (to learn more, refer to chapter 5). In this study, we have also focused on investigating the possible impact of demographic characteristics on EXIT 360° scores. Firstly, there was no gender-related difference in both EXIT 360° global scores. Furthermore, neither age nor education had any effect on the time variable. On the contrary, the EXIT 360° Total Score was influenced by education and age. Firstly, there was a difference in the total score between the low education level (5) and medium-high education groups. Additionally, there was a correlation between age and overall score, with older participants receiving lower ratings. However, considering the joint demographic factor interact on EXIT 360° Total Score, only the effect of the education variable was shown to be significant (with only a tendency to significance for variable "age"). As a result, just like with most neuropsychological tests, it will be necessary to provide a standardization of total scores for age and education.

Additionally, other analyses on the seven subtasks were carried out to examine the performance of participants at each subtask and any correlations between them and neuropsychological tests. These further analyses aimed to determine the 1) potential differences in the complexity of subtasks and 2) EFs evaluated by each. Task 7 appeared to be the most challenging activity (only 64.2% of participants provided the correct answer), followed by Task 6 (66.2%). Except for Task 1, the correlation analysis revealed a growing cognitive burden in the tasks. These findings supported the rationale behind the design of the EXIT 360° activities, according to they were built to increase in terms of cognitive load (number of cognitive components evaluated). However, the difficulty could also be increased by introducing confounding variables (distractors). Indeed, Tasks 5 and 6 assess the same amount of EFs, but Task 6 appeared more difficult in terms of correct answer percentages (90 vs 66.2) due to the addition of confounding variables. Overall, the subtasks displayed a rise in complexity regarding the percentages of right responses, except for Tasks 1 and 5. Task 5 appeared simpler than task 4 (and too simple than task 6); therefore, we have decided to add more distractions to this activity in the following release. On the other hand, Task 1 appeared more complex that Tasks 2 and 3 since it evaluated more EFs, showing a major percentage of incorrect answers (20.8%). As it was developed, it is not possible to shift this task, as Task 1 asks the subject "to plan a strategy to go out at house looking at a map". However, we believe it cannot compromise or penalize the results' global performance or validity since participants must complete the entire task. Data have also shown no correlation between Task 2 and neuropsychological evaluation. This result is not surprising since Task 2 was developed to assess the decision-making that was not measured by the selected tests. As a result, the introduction to the neuropsychological evaluation of a test to measure decision-making ability could demonstrate the capacity of Task 2 to assess this EF. However, a possible explanation could be the "ceiling effect", as all control subjects have performed the task correctly.

Overall, the findings demonstrated that EXIT 360° with its seven subtasks allowed for assessing multiple components of executive functioning. In detail, EXIT 360° emerged as a valuable and promising ecologically valid tool to evaluate a] selected and divided attention (subtasks 3-5–6-7), b] cognitive flexibility (subtasks 1–4–7–6-7), c] inhibition control and interference sensitivity (subtasks 1–4–6-7), d] working memory (subtasks 1–5–6-7), e] planning (subtasks 4-6-7), f] visual search (3–7), g] set-switching (subtask 1–5–6), and h] reasoning (subtask 5-6-7).

Finally, the usability evaluation results supported previous research, demonstrating a "good to excellent" usability score, with over 32% of participants that evaluated EXIT 360° as excellent and 27.6% as best imaginable. Interestingly, data showed no correlation between the total usability score and the Total Score of EXIT 360°. Therefore, the score obtained by the participants to our innovative 360°-based tool is not influenced by the usability level but only by participants' performance (as also highlighted by the correlation between neuropsychological tests and total score).

This chapter summarized the study's results to evaluate the convergent validity of EXIT 360°. This study serves as a further step in validating EXIT 360° as a valid and standardized instrument that exploits the 360° technologies for an ecologically valid assessment of executive functioning. Other studies will be necessary to 1] provide standardization of EXIT 360° Total Score for age and education and 2] evaluate its effectiveness in discriminating between healthy control subjects and patients with executive dysfunctions. In this context, Chapter 5 will detail the main results obtained by the study conducted to evaluate the diagnostic efficacy of EXIT 360° in distinguishing healthy controls from patients with Parkinson's Disease.

5

5. EXIT 360° - Construct Validity and Diagnostic Assessment

Previous Chapters have shown that EXIT 360° appears to be a novel solution within the field of neuropsychological evaluation of executive functionality, showing excellent findings in terms of convergent validity and usability (Borgnis, Baglio, Pedroli, Rossetto, Meloni, et al., 2022). This Chapter will provide the main results of the study conducted to investigate a further important psychometric property of the new tool, namely the construct validity.

Over the years, neurologists and neuropsychologists have increasingly relied on neuropsychological assessment to increase diagnostic accuracy in many neurological disorders to obtain crucial information for developing neurorehabilitation interventions (Braun et al., 2011). Therefore, the “diagnostic efficacy” in discriminating between pathological populations and control groups appears to be a key component in developing and validating a new assessment tool.

In this context, sensitivity and specificity are two criteria usually used to evaluate the construct validity of tests, that is, its ability to identify, among populations, those with the required "character" and those that do not. For example, if the character analyzed is “presence of disease”, the sensitivity and specificity will allow for evaluating the ability of a test to identify healthy and pathological populations. In detail, the sensibility answers the question, “How many of the sick individuals tested were positive?”. At the same time, the specificity responds to the inquiry, “How many of the healthy individuals tested were negative?”. It, therefore, appears evident that the lack of construct validity constitutes a significant limitation in the use of a specific tool since the absence of information on diagnostic specificity and sensitivity in clinical populations makes it impossible to introduce them into clinical practice.

The work done to assess EXIT 360° diagnostic validity in distinguishing individuals with Parkinson’s Disease (PD) from healthy controls is summarized in this chapter.

PD appeared as an ideal clinical population to study since it is well known that executive dysfunction represents a common non-motor symptom in early-stage non-demented PD (Kudlicka et al., 2011 Aarsland et al., 2005). Deficits in attention, planning, set-shifting, dual-task performance, inhibitory control, working memory, and decision-making can be considered the core characteristics of executive dysfunction in PD (Maggio et al., 2018). As a result, patients struggle with many essential goal-directed everyday activities, with substantial adverse implications for daily functioning and quality of life (Barone et al., 2017; Lawson et al., 2016; Leroi et al., 2012). A growing number of longitudinal studies have revealed that early executive dysfunction is predictive of PD conversion in "PD with dementia" (Azuma et al., 2003; Janvin et al., 2005). However, research indicated that traditional standard tests did not appear to be responsive to detect executive deficits in real-world situations (Cipresso et al., 2014). In this context, an early and ecologically valid assessment of the executive profile appears crucial to achieve excellent and effective disease management. EXIT 360° could therefore permit early detection of executive deficits and, consequently, identify patients at risk of developing dementia, providing timely interventions (Cipresso et al., 2014; Serino et al., 2014)

Material and Methods

Participants: Following the sample size calculation reported in the Statistical Analyses section, we involved in the study eighty participants to guarantee optimal statistical power: thirty-six patients with Parkinson's Disease (PwPD group) and 44 healthy controls (HC group) matched for gender, age, and education. PwPD were consecutively recruited by an experienced neurologist at the Parkinson Center of IRCCS Fondazione Don Carlo Gnocchi ONLUS (FDG, Milan, Italy). HC group was recruited among volunteers, family members, and people participating in the public meeting. All participants had to meet inclusion criteria (a) age ranged 18-90 years old; (b) education level ≥5 (primary school); (c) no cognitive impairment as measured by the screening test Montreal Cognitive Assessment (MoCA score ≥15.51), corrected for age and education following Italian normative data (Santangelo et al., 2015); and (d) ability to give a written, signed informed consent. Additionally, PwPD had to satisfy the following inclusion criteria: (a) clinically established or probable Parkinson's disease following Movement Disorder Society (MDS) criteria (Postuma et al., 2015); (b) mild to moderate disease staging with scores < 3 on the Hoehn and Yahr scale. Severe hearing or vision impairments, grave systemic, mental, or other neurological diseases, and overt visual hallucinations or vertigo were considered exclusion criteria.

Procedures: All participants underwent a one-session evaluation at FDG that involved the four steps of the procedure detailed in the "EXIT 360° Validation" of Chapter 2: (a) introduction, (b) neuropsychological evaluation, (c) EXIT 360° session, and (d) usability assessment (Borgnis, Baglio, Pedroli, Rossetto, Meloni, et al., 2021). Briefly:
(a) *Introduction:* the neuropsychologist recorded the main participants' socio-demographic data (i.e., age, education, and sex) and thoroughly explained the study's goals and potential risks before they signed the written informed consent.

(b) *Neuropsychological Evaluation:* A neuropsychologist evaluated the global and executive functioning of participants using standard pencil-and-paper neuropsychological battery:

[a] MoCA test: a sensitive screening tool able to exclude the presence of global cognitive impairment.

[b] Integrated executive functions battery involving Trail Making Test (Reitan, 1992a), phonemic verbal fluency task (F.A.S.) (Novelli et al., 1986), Stroop Test (Stroop, 1935), Digit Span Backward (Monaco et al., 2013), Frontal Assessment Battery (FAB) (Appollonio et al., 2005; Dubois et al., 2000), Attentive Matrices (Spinnler & Tognoni, 1987b) and Progressive Matrices of Raven (PMR) (Caffarra et al., 2003; Raven, 1938) (for a detailed description of administered neuropsychological tests see Chapter 4).

MoCA test was used to verify that participants met the inclusion criteria. On the other hand, conventional tests for EFs were used to confirm the good convergent validity of EXIT 360° and to evaluate its diagnostic effectiveness by comparing it with neuropsychological gold standards.

EXIT 360° Session: Each subject underwent an evaluation with EXIT 360°, preceded by a phase of familiarization with the technological device. Detailed characteristics and administration procedure of EXIT 360° have been described in Chapter 2. Briefly, participants performed everyday subtasks in 360° domestic environments delivered via a smartphone connected to HMD. Participants had to perform all seven subtasks, obtaining one point for a wrong answer or two for a correct one. Overall, EXIT 360° allowed for the collection of Total Score (range 7–14) and Total Time (i.e., time in seconds registered from initial instruction until the participant provided the last correct answer).

Usability Assessment: all participants underwent a usability assessment of the EXIT 360° using the System Usability Scale (SUS) (Brooke, 1986, 1996) to confirm that technological usability has no impact on performance (and, consequently, on results' validity).

Statistical Analysis: The sample size calculation was performed with G power software.

F tests - ANOVA: Fixed effects, omnibus, one-way

Analysis:	A priori: Compute required sample size		
Input:	Effect size f	=	0.369
	α err prob	=	0.05
	Power (1-β err prob)	=	0.85
	Number of groups	=	2
Output:	Noncentrality parameter λ	=	9.2589480
	Critical F	=	3.9862695
	Numerator df	=	1
	Denominator df	=	66
	Total sample size	=	68
	Actual power	**=**	**0.8504415**

The Effect side obtained from the validation study of PIT 360° (Serino et al., 2017) was chosen as a benchmark as it used similar procedures and materials (e.g., 360° settings, experimental procedure, and clinical population involved). According to the sample size calculation, we needed at least 68 subjects to guarantee optimal statistical power (.85).

Descriptive statistics included the frequencies, percentages, median and interquartile range (IQR) for categorical variables and the mean and standard deviation (SD) for continuous ones. The Kolmogorov-Smirnov test was used to determine whether the data distribution was normal. T-tests for independent samples (parametric or non-according to variables) or a chi-squared test for categorical variables were used to compare the demographic (i.e., age, sex, and educational level) and clinical (global cognitive functioning) data of the two groups involved. Additionally, ANOVA 2x2 between subjects was performed to examine differences between the two groups in traditional neuropsychological tests and EXIT 360° scores. The potential relationship between the results of the standard neuropsychological tests and EXIT 360° scores (Total Score and Total Time) was assessed using Pearson's correlation. ROC curves were carried out to evaluate the specificity and sensitivity of each administered test. Regarding system usability, Pearson’s correlation was conducted to examine the relationship between EXIT 360° and usability scores. Moreover, ANOVA between subjects was performed to evaluate any differences in usability between the two groups. All statistical analyses were performed using Jamovi 1.6.7. A *p*-value of <0.05 was considered statistically significant.

Additionally, nonlinear stochastic approximation (i.e., machine learning) methods were used to compare the classification accuracy of traditional neuropsychological assessments versus the EXIT 360° indices for classifying participants into either the "Patients with PD" or "Healthy Controls" groups. We employed different algorithms to compare the predictive value of each one of them to understand which one was the best based on their accuracy. A leave-one-out cross-validation was carried out with the following methods:

1] Logistic Regression, which can provide probabilities and classify new data using continuous and discrete datasets;

2] k-nearest neighbors, which assumes the similarity between the new case/data and available ones and places the new case in the category that is most similar to the available categories;

3] Naive Bayes for distinguishing between the two groups even without any particular assumption for the distribution of the features, and

4] Support Vector Machine to map inputs to higher-dimensional feature spaces that best separate different classes. All these analyses were computed using Python 3.4.

Results

Participants: demographic and clinical characteristics of the whole sample (N=80), divided into two groups (PwPD and HC), are reported in Table 5.1. PwPD (n = 36) had a mean age of 68.7 (SD = 8.22, range = 53-84) and an education age of 13 (IQR = 6, range 5-18); HC had a mean age of 65.5 (SD = 13.8, range = 40-89) and education age of 13 (IQR = 8.50, range 5-18). For both groups, the percentage of females is higher than that of males. Additionally, all subjects showed no cognitive impairment (MoCA_adjusted score ≥ 15.51). No significant differences appeared between groups in demographic characteristics and global cognitive level.

	PwPD N=36	**HC N=44**	**Group Comparison (p-value)**
Age (years, mean (SD))	68.7 (8.22)	65.5 (13.8)	.224
Sex (M: F)	15:21	18:26	.945
Age of education (years, median (IQR))	13 (6)	13 (8.50)	.726
MoCA_adjusted score (mean (SD))	25.8 (2.41)	24.7 (2.72)	.082

Table 5.1: Demographic and clinical characteristics of the whole sample. M = male; F = female; SD = standard deviation; IQR = interquartile range; n = number; MoCA = Montreal Cognitive Assessment; PwPD=patients with Parkinson's Disease; HC= Healthy controls

Traditional Neuropsychological Evaluation: Table 5.2 shows significant differences between the two groups in four paper-and-pencil neuropsychological tests of executive functioning. Specifically, HC achieved higher performance compared to PwPD in FAB score (F (1,78) = 27.81; p<.001) and PMR (F (1,78) = 7.82; p=.007). Additionally, the HC group obtained better results (i.e., less time needed to complete the test) compared to PwPD in TMT-B (F (1,78) = 4.70; p=.033) and TMT-BA (F (1,78) = 5.32; p=.024). Performances on the remaining neuropsychological tests appeared similar between the two groups (p<.05).

	PwPD Mean (SD)	**HC** Mean (SD)	**Group Comparison (p-value)**
Trail Making Test – A	32.68 (16.64)	30.59 (21.91)	.641
Trail Making Test – B	117.28 (105.94)	78.52 (48.32)	**.033**
Trail Making Test – B-A	85.5 (98.11)	49 (34.03)	**.024**
Verbal Fluency	37.81 (11.55)	38 (9.68)	.936
Stroop Test - Errors	0.81 (3.1)	0.45 (0.76)	.463
Stroop Test - Time	19.58 (13.15)	22.77 (13.41)	.289
Digit Span Backward	4.47 (1.09)	4.52 (1.03)	.826
Frontal Assessment Battery	15.71 (1.98)	17.52 (1.03)	**<.001**
Attentive Matrices	47.68 (7.44)	50.34 (6.57)	.094
Progressive Matrices of Raven	30.37 (4.04)	32.49 (2.73)	**.007**

Table 5.2: Comparison of scores at traditional neuropsychological tests. SD = standard deviation; PwPD=patients with Parkinson's Disease; HC= Healthy Controls. (in bold, statistically significant value).

EXIT 360°: Table 5.3 reveals significant differences between the two groups in EXIT 360° Total score (F (1,78) = 70.8; p<.001; $\eta^2 p$=.476) and Total Time (F (1,78) = 52.8; p<.001; $\eta^2 p$=.404). Although all participants were able to complete the test, the HC group obtained a higher Total score than PwPD (mean=12.5±0.95) and completed the test in less time (mean= 484±133.30).

	PwPD Mean (SD)	**HC** Mean (SD)	**Group Comparison (p-value)**
EXIT 360° Total score	10.2±1.46	12.5±0.95	**<.001**
EXIT 360° Total Time	717.4±153.98	484±133.30	**<.001**

Table 5.3: Comparison of scores at EXIT 360°. SD = standard deviation; PwPD=patients with Parkinson's Disease; HC= healthy controls. (in bold, statistically significant value)

Correlation between Neuropsychological Tests and EXIT 360°: Table 5.4 shows significant correlations (Pearson's correlation) between traditional paper-and-pencil neuropsychological tests and the two scores of EXIT 360°. In detail, a significant correlation appeared between the EXIT 360° Total Score and all neuropsychological tests. Additionally, data showed a relationship between EXIT 360° Total Time and various assessment tests, particularly timed ones (Trail Making Test).

	PMR	AM	FAB	VF	DS	TMT-A	TMT-B	TMT-BA	ST-E	ST-T
EXIT-360° Total Score	**.464****	**.271****	**.620#**	**.305***	**.232***	**-.309***	**-.453****	**-.424#**	**-.251***	-.218
EXIT-360° Total Time	**-.333***	-.209	**-.433#**	-.084	-.009	.170	**.477****	**.489#**	.199	.139

Table 5.4: Correlation between EXIT 360° scores and Neuropsychological Assessment. PMR= Progressive Matrices of Raven; AM= Attentive Matrices; FAB= Frontal Assessment Battery; VF= Verbal Fluency; DS= Digit Span; TMT-A= Trail Making Test – part A; TMT-B= Trail Making Test – part B; TMT-BA= Trail Making Test – part B-A; ST-E= Stroop Test – Errors; ST-T= Stroop Test – Time. In bold, statistically significant scores. *p<.05; #p<.001.

Classification of Healthy Controls or Clinical Group: ROC analyses and nonlinear stochastic approximation methods were used to evaluate the accuracy of EXIT 360° scores in discriminating HC and PwPD, showing highly interesting results.

ROC Curve Analysis: The performance of the classifiers was evaluated by carrying out a relative operating characteristic (ROC) analysis. The area under the ROC curve (AUC) provides a single measure of overall prediction accuracy. According to the literature(Mandrekar, 2010), an excellent accuracy value ranges between 0.8 and 0.9. In our study, ROC curves (Figure 5.1) investigated the diagnostic accuracy of EXIT 360°, showing:

(a) EXIT 360° Total Score ≤ 11 could accurately discriminate HC and PwPD groups, with high sensitivity (90.91%) and specificity (77.78%) (AUC=.897 - excellent accuracy value).

(b) EXIT 360° Total Time ≤ 572 could accurately discriminate HC and PwPD groups, with high sensitivity (86.11%) and specificity (86.36%) (AUC=.884 - excellent accuracy value).

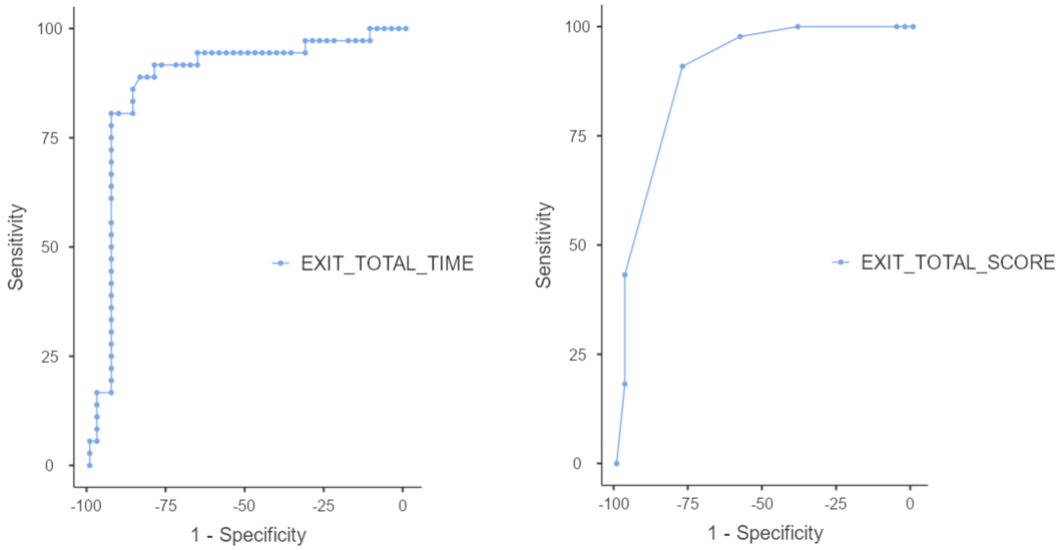
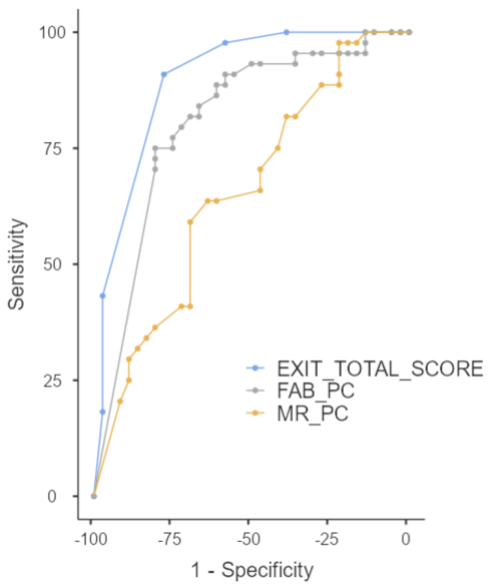


Figure 5.1: ROC Curve – EXIT 360° Total Time and Total Score.

<u>Comparison between ROC curves of EXIT 360° scores and traditional neuropsychological tests</u> (only those that showed a significant difference between the two groups).

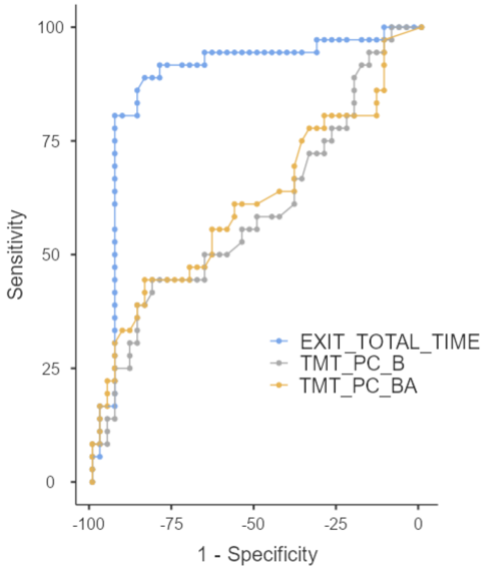
Further analyses showed that EXIT 360° Total Score ≤ 11 can discriminate between HC and PwPD with better overall prediction accuracy, sensitivity, and specificity than MR (DeLong test – $p<.001$) and FAB (DeLong test – $p=.04$) scores (Figure 5.2).



	AUC	Sensitivity	Specificity
EXIT Total Score	.897 (excellent)	90.91%	77.78%
MR	.653 (Not-acceptable)	63.64%	63.89%
FAB	.806 (excellent)	75%	80.56%

Figure 5.2: ROC Curve – Comparison between EXIT 360° Total Score and neuropsychological tests

Similarly, EXIT 360° Total Time≤572 allows for discriminating between HC and PwPD with better overall prediction accuracy, sensitivity, and specificity than TMT-B and TMT-BA (DeLong test – p<.001) scores (Figure 5.3).



	AUC	Sensitivity	Specificity
EXIT Total Score	.884 (excellent)	86.11%	86.36%
TMT-B	.595 (Not-acceptable)	44.44%	81.82%
TMT-BA	.616 (Not-acceptable)	44.44%	84.09%

Figure 5. 3: ROC Curve – Comparison between EXIT 360° Total Time Score and neuropsychological tests.

Nonlinear stochastic approximation methods: confirmed the ROC analyses, showing an excellent accuracy of EXIT 360° scores in discriminating HC and PwPD. Results showed a precision (i.e., the proportion of true positives among all the instances classified as positive) between 61% and 65% for the conventional neuropsychological assessment of EFs (Table 5.5 panel A), while it ranged from 79% to 86% for EXIT 360° (Table 5.5 panel C) and from 80% to 90% for traditional battery and EXIT 360° together (Table 5.5 panel B). Additionally, the value of classification accuracy (CA - i.e., the proportion of the instances that were classified correctly) achieved a range between 79% and 85% for EXIT 360° compared to neuropsychological tests that do not exceed 65%.

METHODS	AUC	CA	F1	PRECISION	RECALL
[A] Traditional Neuropsychological tests					
k-nearest neighbors (kNN)	0.67	0.61	0.61	0.61	0.61
Logistic Regression	0.67	0.61	0.61	0.61	0.61
Naive Bayes	0.68	0.65	0.65	0.65	0.65
Support Vector Machine (SVM)	0.68	0.64	0.63	0.63	0.64
[B] EXIT-360° and Traditional Neuropsychological tests					
k-nearest neighbors (kNN)	0.86	0.80	0.80	0.80	0.80
Logistic Regression	0.93	0.90	0.90	0.90	0.90
Naive Bayes	0.85	0.80	0.80	0.80	0.80
Support Vector Machine (SVM)	0.90	0.81	0.81	0.81	0.81
[C] EXIT-360°					
k-nearest neighbors (kNN)	0.86	0.79	0.79	0.79	0.79
Logistic Regression	0.91	0.85	0.85	0.85	0.85
Naive Bayes	0.91	0.83	0.83	0.83	0.83
Support Vector Machine (SVM)	0.91	0.85	0.85	0.86	0.85

Table 5.5: Leave one out cross-validation (LOOCV) for the traditional neuropsychological tests [A], the indices of EXIT-360° and the traditional neuropsychological tests [B], and the index of EXIT-360° [C]. AUC (Area under the ROC curve) is the area under the classic receiver-operating curve; CA (Classification accuracy) represents the proportion of the examples that were classified correctly; F1 represents the weighted harmonic average of the precision and recall (defined below); Precision represents a proportion of true positives among all the instances classified as positive. In our case, the proportion of conditions correctly identified; Recall represents the proportion of true positives among the positive instances in our data.

Interestingly, all machine learning algorithms showed that the indices from EXIT 360° had a higher capability in predicting PD Group membership compared to traditional neuropsychological tests of executive functioning (Figure 5.4).

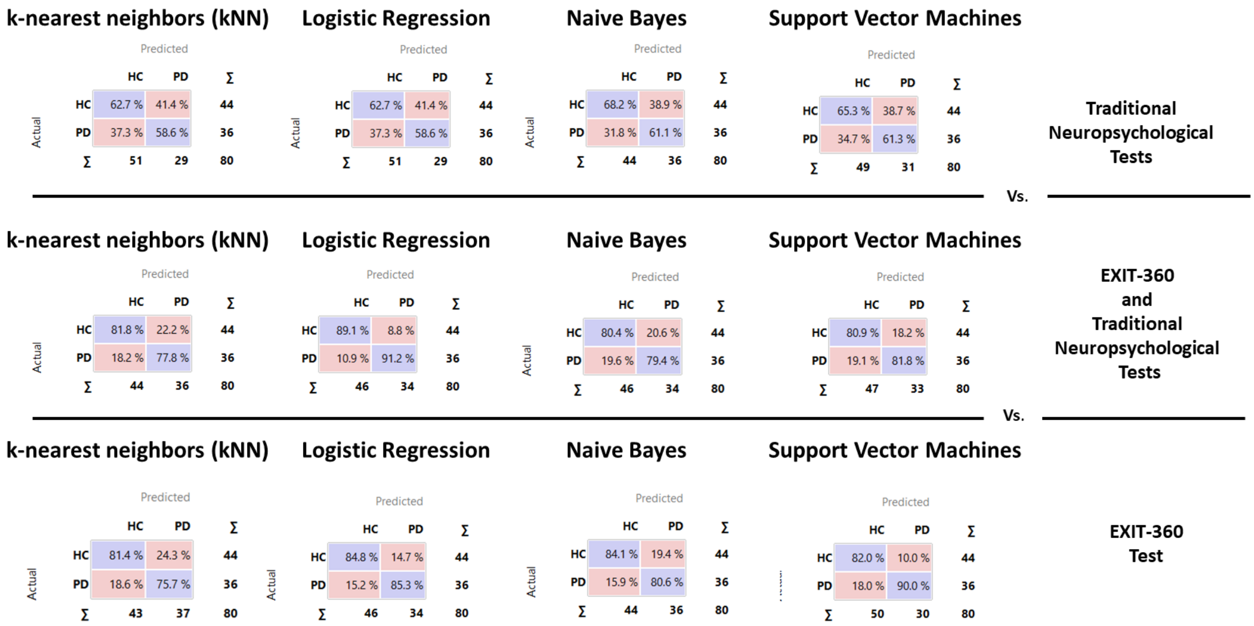


Figure 5.4: Classification of HC or PwPD. The diagonal values (i.e., purple boxes) represent the elements for which the predicted group is equal to the true group, while of-diagonal elements are those that are mislabeled by the classifier. Logistic Regression and Support Vector Machine algorithms demonstrated that EXIT-360° has a higher capability in predicting PD Group membership with respect to traditional neuropsychological tests of executive functioning.

Overall, data showed that integration between neuropsychological tests and EXIT 360° could allow better classification accuracy, with precision ranging between 80% and 90%. Interestingly, the only use of EXIT 360° could allow an excellent classification accuracy, with a precision ≥ 79%.

Additionally, Table 5.6 shows the probability (%) of belonging to the HC group or PD group, considering EXIT 360° Total Time (for a graphical representation, see Figure 5.5).

	HC	PD
≤502	**90.60%**	9.40%
502-524*	**100%**	0%
524-574*	42.9%	**57.1%**
574-600*	**100%**	0%
600-802*	0%	**100%**
>802**	33.3%	**66.7%**
*Included		
**if EXIT 360° Total Score ≤ 10, PD at 100%, if score > 10, HC at 100%.		

Table 5.6: Probability of belonging to the HC group or PD group, considering EXIT 360° Total Time. HC=Healthy Control; PD=Parkinson's Disease

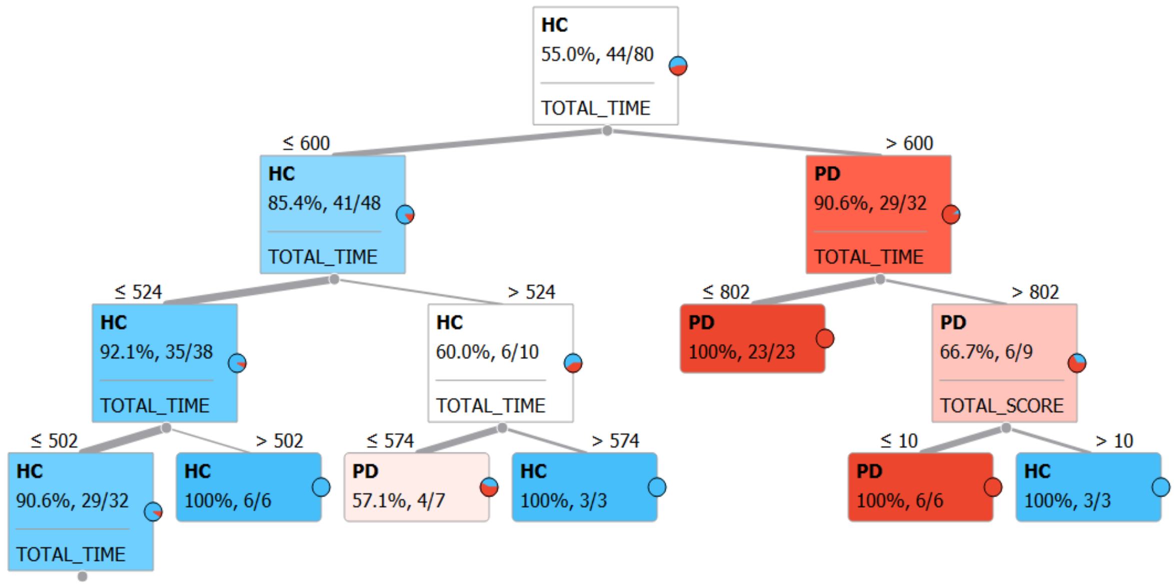


Figure 5.5: Graphic representation of the probability of belonging to the HC group or PD group, considering EXIT 360° Total Time.

Post-hoc Sample size: The post-hoc sample size calculation performed with G power software (see below) showed that our total sample (n=80) allowed us to obtain an excellent statistical power = 0.999.

F tests - ANOVA: Fixed effects, omnibus, one-way

Analysis:	Post hoc: Compute achieved power		
Input:	Effect size f	=	0.9530986
	α err prob	=	0.001
	Total sample size	=	80
	Number of groups	=	2
Output:	Noncentrality parameter λ	=	72.6717553
	Critical F	=	11.6941870
	Numerator df	=	1
	Denominator df	=	78
	Power (1-β err prob)	**=**	**0.9999996**

Usability: The comparison between the PwPD and HC showed the absence of a significant difference in usability score ($F(1,78) = .415$; $p=0.521$). Indeed, PwPD provided a mean score of 77.3±9.30, while HC showed a mean score of 75.7±12.41. Both scores indicate a satisfactory level of usability, according to the scale's score acceptability ranges (cut off = 68) and adjective ratings (included between "good" and "excellent"). Moreover, Pearson's correlation showed no significant relationship between the total usability score and, respectively, EXIT 360° Total Score (p=.711) and EXIT 360° Total Time (p=.560).

Discussion and Conclusion

This Chapter has focused on the main results of a recent work investigating the potentiality of EXIT 360° to integrate the traditional neuropsychological assessment of EFs in PD with a more ecologically valid assessment.

In our work, we compared the performance of PwPD and HC by comparing conventional neuropsychological assessments with the EXIT 360° to evaluate the ability of EXIT 360° in detecting executive deficits. Correlations between neuropsychological tests of EFs and performance on EXIT 360° were also explored. Finally, we investigated the predictive validity of indices obtained from EXIT 360° in discriminating PD patients from HC.

All subjects underwent a neuropsychological assessment of global cognitive functioning, obtaining scores within the normal range. This confirms that patients were in a relatively well-preserved clinical state. However, the neuropsychological evaluation of EFs showed differences between patients and HC in three tests (FAB, TMT, and PMR). Correlation analyses indicated that neuropsychological tests correlate significantly with EXIT 360° scores, supporting a good convergent validity in line with previous results (see Chapter 4). Specifically, the time needed to complete EXIT 360° was significantly correlated with the TMT, tapping visuospatial aspects of executive functioning and motor aspects. Moreover, EXIT 360° Total Score showed a significant correlation with scores of all neuropsychological tests administered, showing the ability of EXIT 360° to evaluate more components of executive functioning, including cognitive flexibility, inhibition control, and sustained and selective attention.

The analyses on EXIT 360° also revealed different performances in PwPD compared to HC, which was consistent with the pattern of results from the traditional neuropsychological assessment. These findings demonstrated the ecological tool's strong sensitivity to executive dysfunction in PD, even in its mild-to-moderate stage, when motor symptoms predominate over cognitive ones. This result assumes considerable importance since executive deficits in the early stage of PD are predictive of the conversion to dementia (Ceravolo et al., 2012; Kudlicka et al., 2011; Levy et al., 2002; Paulwoods & Tröster, 2003), with a negative impact on everyday functioning (Barone et al., 2017; Chan et al., 2008; Leroi et al., 2012). As a result, it is important to identify individuals at a higher risk of developing dementia to develop an early and customized cognitive rehabilitation treatment (Cipresso et al., 2014; Serino et al., 2014). Additionally, it is well known that the most significant issues of conventional neuropsychological tests include their lack of ecological validity and their ability to assess just one component of executive functionality without reflecting an accurate and complex picture of a patient's executive status (Burgess et al., 2006; Chan et al., 2008; Chaytor & Schmitter-Edgecombe, 2003). For this reason, patients with presumed executive deficits can perform similarly to HC on traditional neuropsychological tests while having trouble in everyday life (Burgess et al., 2006). In this context, the technology 360° may be used to provide a new paradigm in which patients are active participants within an ecological virtual world (Parsons, 2015; Riva, 2009), where it is possible to simulate life-like challenges that reproduce everyday situations and, as a result, actual patient's executive status. In this framework, EXIT 360° has proved to be an innovative instrument for detecting executive dysfunction through a function-led approach that combined experimental control with an engaging real-world background.

Our main findings revealed significant differences in EXIT 360° Total Score and EXIT 360° Total Time score between patients and cognitively healthy participants. Particularly, PD patients took longer to complete the test and made more mistakes than HC patients when completing the subtasks of EXIT 360°. Our findings align with a prior study on PwPD, which demonstrated the effectiveness of a VR-based tool, VMET, in assessing executive impairments that had not been fully recognized by the conventional neuropsychological battery (Cipresso et al., 2014). Additionally, according to ROC curve analysis, EXIT 360° scores (accuracy and completion time) have shown a great capacity to distinguish between PwPD and HC. In detail, an EXIT 360° Total Score of ≤11 allows for accurately (AUC=.897) discriminating between PwPD and HC with high sensitivity and specificity. The same outcomes are seen when total time is taken into account, where a score ≤ 572 provides for precise (AUC=.884) differentiation between patients and controls. According to the literature (Mandrekar, 2010), an accuracy value between 0.8 and 0.9 can be considered excellent. Additionally, ROC curve analysis showed that EXIT 360° Total Score and Time Score have better overall prediction accuracy, sensitivity, and specificity than conventional paper-and-pencil neuropsychological evaluation when differentiating between HC and PwPD. These promising findings were confirmed by the higher diagnostic accuracy in machine learning classification of participants to the clinical or non-clinical conditions (when using indices from EXIT 360°) compared to those from neuropsychological assessments. These robust results showed that EXIT 360° is effective at identifying impairments of several components of executive functioning at an early clinical stage of PD. Interestingly, machine learning analyses have also suggested that integration between neuropsychological tests and EXIT 360° could allow better classification accuracy, with precision ranging between 80% and 90%. This result supported the potentiality of EXIT

360° to integrate the traditional neuropsychological assessment of EFs in PD with a more ecologically valid assessment. However, only using EXIT 360° could allow an excellent classification accuracy, with high precision (≥ 79%). EXIT 360° can therefore be seen as an ecological tool highly usable for prompt diagnosis and early patient enrolment in focused rehabilitation.

Overall, in line with other research (Serino et al., 2017), our findings show that 360° technology may play a key role in neuropsychological assessment. Specifically, our results support previous research on the 360° version of the Picture Interpretation Test (PIT) ability to identify executive dysfunction in active visual perception, which is typical of PwPD compared to HC. EXIT 360° might be seen as a development and evolution of PIT 360°, allowing for the evaluation of various executive functioning components in an ecological context. Firstly, the multicomponent dimension appears critical in evaluating EF since it is defined as a complex and heterogeneous construct involving a wide range of cognitive processes and behavioral skills responsible for controlling and regulating actions (e.g., starting and stopping activities or monitoring) (Burgess & Simons, 2005; Chan et al., 2008). This aspect also proves important to PwPD since several studies have shown the presence of several executive impairments in PD, such as planning, attention, working memory, set-shifting, dual-task performance, inhibitory control, and decision making, including social–cognition abilities (Borgnis, Baglio, Pedroli, Rossetto, Uccellatore, et al., 2022; Dahdah et al., 2017; Diamond, 2013; Dirnberger & Jahanshahi, 2013). Secondly, EXIT 360° replicates typical everyday domestic environments such as the kitchen, bedrooms, living room, and landing, enabling an assessment of possible executive impairments in the scenarios most experienced by the subject, with wide implications also in terms of rehabilitation. This feature is peculiar to EXIT 360° because all technological-based instruments for assessing executive functioning in PD have only ever included a small number of real-world settings, especially supermarkets, and never household settings (Borgnis, Baglio, Pedroli, Rossetto, Uccellatore, et al., 2022).

On the advantages of EXIT as an evaluation tool, it is inevitable to mention its usability. Data showed that EXIT 360° can be considered a widely usable tool (obtaining a good to excellent usability score). Additionally, the lack of correlation between usability scores and EXIT 360° indexes supports that EXIT 360° performance is not affected by technological usability issues, in line with a previous study (Borgnis, Baglio, Pedroli, Rossetto, Meloni, et al., 2022). It is a major result since it is well known that technical problems could impact patients' performance, significantly reducing the test results' validity (Armstrong et al., 2013; Parsons & Phillips, 2016). Therefore, this promising result offers a new perspective on the usability of technological instruments in PD neuropsychological assessment, in antithesis to the only study on usability in PD (Pedroli et al., 2013). A critical component in favour of the usability of EXIT 360° is that patients perform the test only by rotating their head, without learning to use technological devices (e.g., joypad) or moving in the environment, with the risk of falling.

While the current study's findings are encouraging, some limitations and future perspectives should also be considered. Firstly, additional research is needed to examine the test-retest reliability of EXIT 360° to fully assess its potential as a new screening tool for EFs. Moreover, it will be crucial to investigate the value of EXIT 360° in detecting executive impairments in other neurological populations, such as Multiple Sclerosis, Mild Cognitive Impairments, and Alzheimer's Disease, that could show executive impairments. Finally, it will be of fundamental importance to design, develop and validate a parallel form of EXIT 360° to make possible a short-term revaluation in a rehabilitation process.

In conclusion, this study offers clear evidence that a more ecologically valid assessment of executive functioning will be more likely to pick up on subtle executive deficits in PD patients. In detail, EXIT 360° captures early executive dysfunctions of PD patients with better diagnostic sensibility and specificity than the traditional paper-and-pencil neuropsychological battery. We believe that this innovative 360°-based tool, which is simply used in clinical settings, has the potential to radically transform patients' and clinicians' evaluation experience. Firstly, since EXIT 360° only lasts at most 15 minutes, the evaluation timeframes for executive functionality will be drastically shortened. Additionally, neurologists and neuropsychologists can obtain ecologically valid multicomponent assessments of executive functioning in PD, gathering data on the patients' actual executive status. The ecological evaluation will enable clinicians to tailor rehabilitation to real everyday subjects' needs. As previously said, early treatment of executive dysfunction in early-stage non-demented PD could minimize the severity and the impact of this key clinical non-motor symptom, improving the patient's daily functioning and quality of life (Kudlicka et al., 2011; Maggio et al., 2018). Interestingly, as EXIT 360° was designed, a streaming platform might potentially use it to carry out remote assessments, overcoming the social distancing limits.

6

6. Discussion and Conclusion

This thesis has focused on the development process of EXIT 360° from its concept to validation as a sensitive new-brand tool for an ecologically valid and multicomponent assessment of executive functioning (Borgnis, Baglio, Pedroli, Rossetto, Riva, et al., 2021). EXIT 360° was conceived and designed to be an original 360°-based instrument involving participants in a new "game for health" in which they are immersed in 360° domestic environments delivered via a smartphone and conventional head-mounted display. In these settings, subjects must perform and complete seven everyday subtasks of increasing complexity designed to assess several aspects of executive functioning simultaneously and quickly. EXIT 360° exploited the advances in 360° technology that recently emerged as a valuable alternative approach for neuropsychological assessment, able to provide an ecological evaluation of executive functions that can predict real-world performance (Negro Cousa et al., 2019; Realdon et al., 2019; Serino et al., 2017). It is well known that executive function is a complex and heterogeneous construct involving various cognitive processes and behavioural skills responsible for controlling and regulating actions and performing complex or non-routine tasks (M. K. Alderman, 2013; Chan et al., 2008). Therefore, executive dysfunction, typical of several psychiatric and neurologic conditions, constitutes a significant global health challenge due to its high impact on personal independence (e.g., preparing meals, managing money, shopping, using a telephone), ability to work, educational success and social relationships, with obvious repercussions on the quality of life (Chan et al., 2008; Diamond, 2013). The executive functions are traditionally evaluated using paper-and-pencil neuropsychological tests that guarantee highly standardized procedures and scores that make them valid and reliable. However, the literature shows the inability of these traditional tests to predict the complexity of executive functioning in real-life settings (Burgess et al., 2006). Indeed, patients with presumed executive impairments can perform similarly to HC on conventional neuropsychological tests while having trouble in everyday life.

As a result, identifying early innovative methods to detect executive impairments appears a priority for planning timely rehabilitation programs.

Over the years, several studies have shown the feasibility and acceptability of VR-based tools in the early assessment of executive functioning in healthy controls and many psychiatric and neurologic pathologies (Aubin et al., 2018; Camacho-Conde & Climent, 2020; Cipresso et al., 2014; Dahdah et al., 2017). However, a recent systematic review (2022) has shown several psychometric issues in the available VR-based assessment tools for executive functioning due to limited studies on construct validity, convergent validity, discriminant validity, usability, and test re-test reliability (Borgnis, Baglio, Pedroli, Rossetto, Uccellatore, et al., 2022).

This thesis offered a comprehensive overview of the studies conducted to evaluate the psychometric properties of EXIT 360°, namely usability, convergent validity and construct validity, to answer three questions: [1] "Will the EXIT 360 ° technology be usable by both the people with low technological expertise, elderly and patients with cognitive dysfunction?”; [2] "Will EXIT 360° actually evaluate multiple components of executive functioning?"; [3] "Will EXIT 360° be able to discriminate between patients and healthy subjects with good diagnostic accuracy?". As widely said in previous Chapters, Parkinson’s Disease (PD) was chosen as the clinical condition since it is well known that executive dysfunction represents a common non-motor symptom in early-stage non-demented PD (Kudlicka et al., 2011 Aarsland et al., 2005), with substantial adverse implications for daily functioning and quality of life.

[1] *Will the EXIT 360 ° technology be usable by the people with low technological expertise, the elderly, and patients with cognitive dysfunction?*”

In Chapter 3, it has been widely demonstrated that EXIT 360° can be considered a highly usable instrument by people with low technological expertise, elderly subjects and patients with PD (PwPD) who also claimed to have had an excellent experience using it (Borgnis, Baglio, Pedroli, Rossetto, Isernia, et al., 2021; Borgnis, Baglio, Pedroli, Rossetto, Meloni, et al., 2022). The importance of evaluating usability and user experience quality has grown over the years, becoming a prerequisite in creating, developing, and validating digital content (Aubin et al., 2018; Schultheis & Rizzo, 2001). All participants considered EXIT 360° as a usable and easy-to-learn technological tool suitable for their skill levels (Borgnis, Baglio, Pedroli, Rossetto, Meloni, et al., 2022). In other words, the new tool demonstrated high effectiveness (i.e., good ability for users to achieve goals), efficiency (i.e., few users' efforts to achieve the goal), and satisfaction (i.e., good users' thoughts about their interaction with the system) (Iso, 1998). Therefore, it is conceivable to argue that any subject's low performance will not depend on technological problems. The excellent usability can be considered a key characteristic of EXIT 360° since it is well known that the technological issue could impact patients' performance, significantly decreasing the test results' validity (Armstrong et al., 2013; Parsons & Phillips, 2016). The encouraging results allowed for hypothesizing that our system does not need further adjustments to be used by PwPD. These studies offer a new perspective on the usability of technological instruments in PD neuropsychological assessment, in antithesis to the only prior usability study in PD (Pedroli et al., 2013), where patients showed many usability difficulties and the tool required more than an improvement to be considered usable. A critical component in favor of EXIT 360° usability could be that patients perform the test only by moving their head, without learning to use technological devices (e.g., joypad) or moving in the environment, with the risk of falling.

Interestingly, a deepened analysis of any impact of technological expertise on using EXIT 360° showed encouraging results. Despite the low familiarity and competence with the technologies, all subjects were able to complete the test and evaluate it as usable and easy-to-learn. As a result, EXIT 360° appeared as a viable tool that clinicians could use even with patients without prior exposure to technology. Regarding age's impact on performance, older adults completed the assessment with only some small additional suggestions. In addition, these studies showed that participants supported an optimal experience using EXIT 360°. All individuals completed EXIT 360° without experiencing any significant adverse effects. It is a significant result since the feeling of cybersickness can lead to unpleasant experiences for users, impacting their performance and significantly decreasing the test results' validity. Patients and healthy controls supported a positive overall impression of EXIT 360° describing it as clear, pleasant, engaging, funny and not boring. EXIT 360° can be considered an innovative, efficient, straightforward, and fast technological solution with great hedonic quality regarding stimulation (challenging, exciting, and engaging) and novelty. The good spatial presence and engagement, excellent ecological validity, and irrelevant adverse effects appeared as crucial features of EXIT 360. The high levels of engagement, enjoyment, and attractiveness of EXIT 360° could increase users' motivation and participation while reducing the fear and anxiety typically associated with neuropsychological testing. Indeed, patients reported they had negative expectations due to customarily drawn-out and complicated evaluations but were pleasantly surprised.

[2] "*Will EXIT 360° actually evaluate multiple components of executive functioning?*"

Chapter 4 reported clear evidence of the good convergent validity of EXIT 360°, comparing it with traditional paper-and-pencil tests, considered the “gold standard” in neuropsychological assessment. It is well known that a strong positive correlation between a new tool and other instruments designed on the same construct is evidence of the convergent validity of the new test (Chin et al., 2014). Therefore, the high correlation found between all indexes of EXIT 360° and scores of the other standardized instruments that evaluate the same construct (i.e., executive functioning) allows for supporting the high level of convergent validity of the new tool. In detail, data revealed a significant correlation between the EXIT 360° Total Score and all neuropsychological tests for executive functioning. An interesting and promising association was also found between EXIT 360° Total Time and timed neuropsychological tests, like the Trail Making Test, Stroop Test, and Attentive Matrices. This study has clearly added an important piece to the development process of EXIT 360° since it can be considered an original and challenging tool able to simultaneously evaluate real-life impairments in several components of executive functioning, namely selective and divided attention, cognitive flexibility, set-shifting, working memory, reasoning, inhibition, and planning. The multicomponent dimension appears critical in evaluating EF since it is a complex and heterogeneous construct involving a wide range of cognitive processes and behavioural skills responsible for many everyday activities (Chan et al., 2008). For example, this aspect could be important in the evaluation of PwPD since several studies have shown the presence of several executive impairments in PD (Borgnis, Baglio, Pedroli, Rossetto, Uccellatore, et al., 2022; Dahdah et al., 2017; Diamond, 2013; Dirnberger & Jahanshahi, 2013).

Interestingly, a deeper analysis confirmed that an evaluation with EXIT 360° did not require a long administration time; indeed, participants took, on average, about 8 minutes to complete the whole task. It is very important since EXIT 360° can allow an evaluation of multiple components of executive functions in a short time, overcoming too long neuropsychological evaluation. Moreover, we showed that all demographic characteristics did not impact the time variable. On the contrary, considering the joint demographic factor interact on EXIT 360° Total Score, an effect of the education variable appeared (with only a tendency to significance for variable "age"). As a result, just like with most neuropsychological tests, it would be necessary to provide a standardization of total scores for age and education.

[3] "*Will EXIT 360° be able to discriminate between patients and healthy subjects with good diagnostic accuracy*?".

The last study, reported in Chapter 5, has successfully investigated the diagnostic efficacy of EXIT 360° in distinguishing between healthy controls and PwPD. Over the years, neurologists and neuropsychologists have increasingly relied on neuropsychological evaluation to improve diagnostic accuracy in many neurological disorders to obtain crucial data for creating neurorehabilitation interventions tailored to patients' needs (Braun et al., 2011). Therefore, the "diagnostic efficacy" in discriminating between pathological populations and control groups appears to be a key component in developing and validating a new assessment tool. Results showed significant differences in EXIT 360° Total Score and Total Time between patients and cognitively healthy participants. Findings revealed that PwPD made significantly more errors in completing EXIT 360° and took longer to conclude the test. These findings demonstrated the ecological tool's strong sensitivity to executive dysfunction in PD, even in its mild-to-moderate stage, when motor symptoms predominate over cognitive ones. These findings align with a prior study on PD, which demonstrated the effectiveness of a VR-based tool in assessing executive impairments that had not been fully recognized by the conventional neuropsychological battery (Cipresso et al., 2014). Classification analysis confirmed a great potential of the EXIT 360° for distinguishing between PwPD and controls. In detail, an EXIT 360° Total score = 12 and EXIT 360° Total Time = 573 seconds (about 10 minutes) can be considered accurate cuts off for discriminating between PwPD and controls with high sensitivity and specificity. Moreover, indices from EXIT 360° showed higher diagnostic accuracy in predicting PD group membership compared to traditional neuropsychological tests. It is well known that the most significant issues of conventional neuropsychological tests include their lack of ecological validity and their ability to assess just one component of executive functionality without reflecting an accurate and complex picture of a patient's executive status (Burgess et al., 2006; Chan et al., 2008; Chaytor & Schmitter-Edgecombe, 2003). For this reason, patients with possible executive difficulties can perform similarly to controls on traditional neuropsychological tests while having trouble in everyday life (Burgess et al., 2006). In this context, EXIT 360° provides a new paradigm in which patients are active participants within an ecological virtual world (Parsons, 2015; Riva, 2009), where it is possible to simulate life-like challenges that reproduce everyday situations and, as a result, actual patient's executive status. Machine learning analyses have also suggested that only using EXIT 360° could allow an excellent classification accuracy, with high precision (≥ 79%). EXIT 360°

can therefore be seen as an ecological tool highly usable for prompt diagnosis and early patient enrolment in focused rehabilitation. This aspect assumes considerable importance, for example, in PD, since executive deficits in the early stage of PD are predictive of the conversion to dementia (Ceravolo et al., 2012; Kudlicka et al., 2011; Levy et al., 2002), with a negative impact on everyday functioning (Barone et al., 2017; Chan et al., 2008; Leroi et al., 2012). As a result, identifying individuals at a higher risk of developing dementia could allow for developing an early and customized cognitive rehabilitation treatment (Cipresso et al., 2014; Serino et al., 2014).

In conclusion, this thesis offers clear evidence that EXIT 360° must be considered a valuable and usable innovative tool for an ecologically valid multicomponent evaluation of executive functioning. EXIT 360° is able to capture early executive dysfunctions of a specific clinical condition with better diagnostic sensibility and specificity than the traditional paper-and-pencil neuropsychological battery. I strongly believe that this innovative 360°-based tool, which is simply usable in clinical settings, has the potential to radically transform patients' and clinicians' evaluation experience. Firstly, since EXIT 360° only lasts at most 15 minutes, the evaluation timeframes for executive functionality will be drastically shortened. Additionally, neurologists and neuropsychologists can obtain ecologically valid multicomponent assessments of executive functioning in PD, gathering data on the patients' actual executive status. The ecological evaluation will enable clinicians to tailor rehabilitation to real everyday subjects’ needs.

Future Perspective: These highly encouraging results pave the way for new and interesting future studies: 1] assessment of EXIT 360° test-retest and inter-rater reliability to deepen its potential as a new screening tool for EFs; 2] investigation of EXIT 360° capacity in detecting executive impairments in other neurological populations that show executive impairments, such as Multiple Sclerosis, Mild Cognitive Impairments and Alzheimer’s Disease; 3] standardization of EXIT 360° Total Score for age and education, as pointed out by convergent validity study; 4] development and validation of a parallel form of EXIT 360° to make possible a short-term revaluation in a rehabilitation process.

Acknowledgements

Questa tesi è il risultato di un grande lavoro di squadra che ha visto in questi ultimi tre anni numerose persone prestare impegno, sacrificio e instancabile supporto affinché questo lavoro si realizzasse.

Vorrei innanzitutto ringraziare il mio supervisor di dottorato Prof. Pietro Cipresso per il supporto, i suoi indispensabili consigli, le conoscenze trasmesse e la sua presenza costante, elementi fondamentali nella riuscita del mio percorso.

Un sentito ringraziamento va inoltre alla Dott.ssa Francesca Baglio per avermi accompagnata lungo questo cammino, offrendomi continue occasioni di apprendimento, contribuendo così alla mia crescita personale e professionale.

Un grazie di cuore alla mia collega, ma soprattutto amica, Federica Rossetto per avermi costantemente supportata e sopportata durante questo mio percorso di vita.

Grazie ai miei colleghi di dottorato, in particolare a Vanessa con cui ho condiviso questa avventura per avermi sempre incoraggiata. Grazie, inoltre, ai miei colleghi del CADiTeR, per farmi sentire ogni giorno parte di un gruppo che sa cosa vuol dire fare squadra e lavorare insieme per uno stesso obiettivo.

Un enorme ringraziamento è dedicato al mio compagno di vita Andrea per essere stato in grado di stare al mio fianco, supportandomi e condividendo con me la fatica di questo lungo viaggio. Ringrazio infinitamente i miei genitori e la mia famiglia per avermi sempre sostenuta, appoggiando ogni mia decisione.

Vorrei inoltre ringraziare tutte le persone che hanno partecipato alla mia ricerca, dedicandomi parte del loro tempo.

Infine, dedico questa tesi anche a me stessa e alla mia tenacia che mi ha permesso di raggiungere questo traguardo, conscia di essere profondamente cresciuta a livello personale e professionale.

References

Aarsland, D., Brønnick, K., Ehrt, U., de Deyn, P. P., Tekin, S., Emre, M., & Cummings, J. L. (2007). Neuropsychiatric symptoms in patients with Parkinson's disease and dementia: frequency, profile and associated care giver stress. *Journal of Neurology, Neurosurgery & Psychiatry*, *78*(1), 36–42.

Aarsland, D., Creese, B., Politis, M., Chaudhuri, K. R., Weintraub, D., & Ballard, C. (2017). Cognitive decline in Parkinson disease. *Nature Reviews Neurology*, *13*(4), 217–231.

Aarsland, D., Zaccai, J., & Brayne, C. (2005). A systematic review of prevalence studies of dementia in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *20*(10), 1255–1263.

Alderman, M. K. (2013). *Motivation for achievement: Possibilities for teaching and learning*. Routledge.

Alderman, N., Burgess, P. W., Knight, C., & Henman, C. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *J Int Neuropsych Soc*, *9*(1), 31–44.

Allamanno, N., della Sala, S., Laiacona, M., Pasetti, C., & Spinnler, H. (1987). Problem solving ability in aging and dementia: Normative data on a non-verbal test. *The Italian Journal of Neurological Sciences*, *8*(2), 111–119.

Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: a meta-analytic review. *Neuropsychology Review*, *16*(1), 17–42.

Appollonio, I., Leone, M., Isella, V., Piamarta, F., Consoli, T., Villa, M. L., Forapani, E., Russo, A., & Nichelli, P. (2005). The Frontal Assessment Battery (FAB): normative values in an Italian population sample. *Neurological Sciences*, *26*(2), 108–116.

Armstrong, C. M., Reger, G. M., Edwards, J., Rizzo, A. A., Courtney, C. G., & Parsons, T. D. (2013). Validity of the Virtual Reality Stroop Task (VRST) in active duty military. *Journal of Clinical and Experimental Neuropsychology*, *35*(2), 113–123. https://doi.org/10.1080/13803395.2012.740002

Aubin, G., Béliveau, M. F., & Klinger, E. (2018). An exploration of the ecological validity of the Virtual Action Planning–Supermarket (VAP-S) with people with schizophrenia. *Neuropsychological Rehabilitation*, *28*(5), 689–708. https://doi.org/10.1080/09602011.2015.1074083

Aubin, G., Stip, E., Gélinas, I., Rainville, C., & Chapparo, C. (2009). Daily activities, cognition and community functioning in persons with schizophrenia. *Schizophrenia Research*, *107*(2–3), 313–318.

Azuma, T., Cruz, R. F., Bayles, K. A., Tomoeda, C. K., & Montgomery Jr, E. B. (2003). A longitudinal study of neuropsychological change in individuals with Parkinson's disease. *International Journal of Geriatric Psychiatry*, *18*(11), 1043–1049.

Baddeley, A., della Sala, S., Papagno, C., & Spinnler, H. (1997). Dual-task performance in dysexecutive and nondysexecutive patients with a frontal lesion. *Neuropsychology*, *11*(2), 187.

Baddeley, A., & Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition*, *7*(2), 212–230.

Bailey, P. E., Henry, J. D., Rendell, P. G., Phillips, L. H., & Kliegel, M. (2010). Dismantling the "age–prospective memory paradox": The classic laboratory paradigm simulated in a naturalistic setting. *Quarterly Journal of Experimental Psychology*, *63*(4), 646–652.

Barker, L. A., Andrade, J., & Romanowski, C. A. J. (2004). Impaired implicit cognition with intact executive function after extensive bilateral prefrontal pathology: A case study. *Neurocase*, *10*(3), 233–248.

Barone, P., Erro, R., & Picillo, M. (2017). Quality of life and nonmotor symptoms in Parkinson's disease. In *International review of neurobiology* (Vol. 133, pp. 499–516). Elsevier.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 1–3.

Benton, A. L. (1991). The prefrontal region: Its early history. *Frontal Lobe Function and Dysfunction*, *19*.

Bialkova, S., & Dickhoff, B. (2019). Encouraging Rehabilitation Trials: The Potential of 360 Immersive Instruction Videos. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 1443–1447.

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, *12*(12), 752–762.

Borgnis, F., Baglio, F., Pedroli, E., Rossetto, F., Isernia, S., Uccellatore, L., Riva, G., & Cipresso, P. (2021). EXecutive-Functions Innovative Tool (EXIT 360°): A Usability and User Experience Study of an Original 360°-Based Assessment Instrument. *Sensors*, *21*(17), 5867.

Borgnis, F., Baglio, F., Pedroli, E., Rossetto, F., Meloni, M., Riva, G., & Cipresso, P. (2021). EXIT 360—executive-functions innovative tool 360—a simple and effective way to study executive functions in parkinson's disease by using 360 videos. *Applied Sciences*, *11*(15), 6791.

Borgnis, F., Baglio, F., Pedroli, E., Rossetto, F., Meloni, M., Riva, G., & Cipresso, P. (2022). A Psychometric Tool for Evaluating Executive Functions in Parkinson's Disease. *Journal of Clinical Medicine*, *11*(5), 1153.

Borgnis, F., Baglio, F., Pedroli, E., Rossetto, F., Riva, G., & Cipresso, P. (2021). A Simple and Effective Way to Study Executive Functions by Using 360° Videos. *Frontiers in Neuroscience*, *15*(April), 1–9. https://doi.org/10.3389/fnins.2021.622095

Borgnis, F., Baglio, F., Pedroli, E., Rossetto, F., Uccellatore, L., Oliveira, J. A. G., Riva, G., & Cipresso, P. (2022). Available Virtual Reality-Based Tools for Executive Functions: A Systematic Review. *Frontiers in Psychology*, *13*.

Braun, M., Tupper, D., Kaufmann, P., McCrea, M., Postal, K., Westerveld, M., Wills, K., & Deer, T. (2011). Neuropsychological assessment: A valuable tool in the diagnosis and management of neurological, neurodevelopmental, medical, and psychiatric disorders. *Cognitive and Behavioral Neurology*, *24*(3), 107–114.

Brooke, J. (1986). System usability scale (SUS): a quick-and-dirty method of system evaluation user information. *Reading, UK: Digital Equipment Co Ltd*, *43*.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4–7.

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L. M. A., Dawson, D. R., Anderson, N. D., Gilbert, S. J., Dumontheil, I., & 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

Burgess, P. W., & Simons, J. S. (2005). 18 Theories of frontal lobe executive function: clinical applications. *The Effectiveness of Rehabilitation for Cognitive Deficits*, 211.

Burke, S. N., & Barnes, C. A. (2006). Neural plasticity in the ageing brain. *Nature Reviews Neuroscience*, *7*(1), 30–40.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*(6), 215–222.

Buter, T. C., van den Hout, A., Matthews, F. E., Larsen, J. P., Brayne, C., & Aarsland, D. (2008). Dementia and survival in Parkinson disease: a 12-year population study. *Neurology*, *70*(13), 1017–1022.

Caffarra, P., Vezzadini, G., Zonato, F., Copelli, S., & Venneri, A. (2003). A normative study of a shorter version of Raven's progressive matrices 1938. *Neurological Sciences*, *24*(5), 336–339.

Camacho-Conde, J. A., & Climent, G. (2020). Attentional profile of adolescents with ADHD in virtual-reality dual execution tasks: A pilot study. *Applied Neuropsychology: Child*, 1–10.

Campbell, Z., Zakzanis, K. K., Jovanovski, D., Joordens, S., Mraz, R., & Graham, S. J. (2009). Utilizing virtual reality to improve the ecological validity of clinical neuropsychology: an FMRI case study elucidating the neural basis of planning by comparing the Tower of London with a three-dimensional navigation task. *Applied Neuropsychology*, *16*(4), 295–306. https://doi.org/10.1080/09084280903297891

Ceravolo, R., Pagni, C., Tognoni, G., & Bonuccelli, U. (2012). The epidemiology and clinical manifestations of dysexecutive syndrome in Parkinson's disease. *Frontiers in Neurology, 3*, 159.

Chan, R. C. K., Shum, D., Toulopoulou, T., & 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

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review, 13*(4), 181–197.

Cherniack, E. P. (2011). Not just fun and games: applications of virtual reality in the identification and rehabilitation of cognitive disorders of the elderly. *Disability and Rehabilitation: Assistive Technology*, *6*(4), 283–289.

Chin, C. L., & Yao, G. (2014). Convergent validity. *Encyclopedia of Quality of Life and Well-Being Research*, *1*(1).

Cipresso, P., Albani, G., Serino, S., Pedroli, E., Pallavicini, F., Mauro, A., & 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*(DEC), 1–11. https://doi.org/10.3389/fnbeh.2014.00405

Cipresso, P., la Paglia, F., la Cascia, C., Riva, G., Albani, G., & la Barbera, D. (2013). Break in volition: A virtual reality study in patients with obsessive-compulsive disorder. *Experimental Brain Research*, *229*(3), 443–449. https://doi.org/10.1007/s00221-013-3471-y

Cipresso, P., Meriggi, P., Carelli, L., Solca, F., Poletti, B., Lulé, D., Ludolph, A. C., Silani, V., & Riva, G. (2013). Cognitive assessment of executive functions using brain computer interface and eye-tracking. *EAI Endorsed Trans. Ambient Syst.*, *1*(2), e4.

Climent, G., Banterla, F., & Iriarte, Y. (2010). Virtual reality, technologies and behavioural assessment. *AULA, Ecological Evaluation of Attentional Processes*, 19–28.

Cowan, N., Saults, J. S., & Blume, C. L. (2014). Central and peripheral components of working memory storage. *Journal of Experimental Psychology: General*, *143*(5), 1806.

Crawford, J. R. (1998). Introduction to the assessment of attention and executive functioning. *Neuropsychological Rehabilitation*, *8*(3), 209–211.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Dahdah, M. N., Bennett, M., Prajapati, P., Parsons, T. D., Sullivan, E., & Driver, S. (2017). Application of virtual environments in a multi-disciplinary day neurorehabilitation program to improve executive functioning using the Stroop task. *NeuroRehabilitation*, *41*(4), 721–734. https://doi.org/10.3233/NRE-172183

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1346), 1413–1420.

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality*, *62*(1), 119–142.

Denmark, T., Fish, J., Jansari, A., Tailor, J., Ashkan, K., & Morris, R. (2019). Using virtual reality to investigate multitasking ability in individuals with frontal lobe lesions. *Neuropsychological Rehabilitation*, *29*(5), 767–788.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64,* 135–168.

Dirnberger, G., & Jahanshahi, M. (2013). Executive dysfunction in P arkinson's disease: A review. *Journal of Neuropsychology*, *7*(2), 193–224.

Dubois, B., Slachevsky, A., Litvan, I., & Pillon, B. (2000). The FAB: a frontal assessment battery at bedside. *Neurology*, *55*(11), 1621–1626.

Duncan, J. (1986). Disorganisation of behaviour after frontal lobe damage. *Cognitive Neuropsychology*, *3*(3), 271–290.

Elkind, J. S., Rubin, E., Rosenthal, S., Skoff, B., & Prather, P. (2001). A simulated reality scenario compared with the computerized Wisconsin Card Sorting Test: An analysis of preliminary results. *Cyberpsychology and Behavior*, *4*(4), 489–496. https://doi.org/10.1089/109493101750527042

Engeser, S., Rheinberg, F., Vollmeyer, R., & Bischoff, J. (2005). Motivation, flow-Erleben und Lernleistung in universitären Lernsettings. *Zeitschrift Für Pädagogische Psychologie*, *19*(3), 159–172.

Fang, C., Lv, L., Mao, S., Dong, H., & Liu, B. (2020). Cognition Deficits in Parkinson's Disease: Mechanisms and Treatment. *Parkinson's Disease*, *2020*.

Fengler, S., Liepelt-Scarfone, I., Brockmann, K., Schäffer, E., Berg, D., & Kalbe, E. (2017). Cognitive changes in prodromal Parkinson's disease: a review. *Movement Disorders*, *32*(12), 1655–1666.

Fortin, S., Godbout, L., & Braun, C. M. J. (2003). Cognitive structure of executive deficits in frontally lesioned head trauma patients performing activities of daily living. *Cortex*, *39*(2), 273–291.

Gilboa, Y., Jansari, A., Kerrouche, B., Uçak, E., Tiberghien, A., Benkhaled, O., Aligon, D., Mariller, A., Verdier, V., & Mintegui, A. (2019). Assessment of executive functions in children and adolescents with acquired brain injury (ABI) using a novel complex multi-tasking computerised task: The Jansari assessment of Executive Functions for Children (JEF-C©). *Neuropsychological Rehabilitation*, *29*(9), 1359–1382.

Gitlin, L. N., Corcoran, M., Winter, L., Boyce, A., & Hauck, W. W. (2001). A randomized, controlled trial of a home environmental intervention: effect on efficacy and upset in caregivers and on daily function of persons with dementia. *The Gerontologist*, *41*(1), 4–14.

Godefroy, O. (2003). Frontal syndrome and disorders of executive functions. *Journal of Neurology*, *250*(1), 1–6.

Godefroy, O., Martinaud, O., Narme, P., Joseph, P. A., Mosca, C., Lhommée, E., Meulemans, T., Czernecki, V., Bertola, C., Labauge, P., Verny, M., Bellmann, A., Azouvi, P., Bindschaedler, C., Bretault, E., Boutoleau-Bretonniere, C., Robert, P., Lenoir, H., Krier, M., & Roussel, M. (2018). Dysexecutive disorders and their diagnosis: A position paper. *Cortex*, *109*, 322–335. https://doi.org/10.1016/j.cortex.2018.09.026

Goel, V., Grafman, J., Tajik, J., Gana, S., & Danto, D. (1997). A study of the performance of patients with frontal lobe lesions in a financial planning task. *Brain: A Journal of Neurology*, *120*(10), 1805–1822.

Golden, C. J., & Freshwater, S. M. (1978). *Stroop color and word test*.

Goldstein, G. (1996). *Functional considerations in neuropsychology.*

Green, M. F. (1996). What are the functional consequences of neurocognitive deficits in schizophrenia? *The American Journal of Psychiatry*.

Green, M. F. (2006). Cognitive impairment and functional outcome in schizophrenia and bipolar disorder. *The Journal of Clinical Psychiatry*, *67*(10), e12.

Harvey, P. D. (2012). Clinical applications of neuropsychological assessment. *Dialogues in Clinical Neuroscience*, *14*(1), 91.

Hely, M. A., Reid, W. G. J., Adena, M. A., Halliday, G. M., & Morris, J. G. L. (2008). The Sydney multicenter study of Parkinson's disease: the inevitability of dementia at 20 years. *Movement Disorders*, *23*(6), 837–844.

Hørlyck, L. D., Obenhausen, K., Jansari, A., Ullum, H., & Miskowiak, K. W. (2021). Virtual reality assessment of daily life executive functions in mood disorders: associations with neuropsychological and functional measures. *Journal of Affective Disorders*, *280*, 478–487.

ISO, B., & STANDARD, B. (2010). *Ergonomics of human-system interaction*.

Iso, W. (1998). 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs). *The International Organization for Standardization*, *45*(9).

Jansari, A. S., Devlin, A., Agnew, R., Akesson, K., Murphy, L., & Leadbetter, T. (2014). Ecological assessment of executive functions: A new virtual reality paradigm. *Brain Impairment*, *15*(2), 71–87. https://doi.org/10.1017/BrImp.2014.14

Janvin, C. C., Aarsland, D., & Larsen, J. P. (2005). Cognitive predictors of dementia in Parkinson's disease: a community-based, 4-year longitudinal study. *Journal of Geriatric Psychiatry and Neurology*, *18*(3), 149–154.

Josman, N., Schenirderman, A. E., Klinger, E., & Shevil, E. (2009). Using virtual reality to evaluate executive functioning among persons with schizophrenia: A validity study. *Schizophrenia Research*, *115*(2–3), 270–277. https://doi.org/10.1016/j.schres.2009.09.015

Júlio, F., Malhão, A., Pedrosa, F., Gonçalves, H., Simões, M., Simões, M. R., van Asselen, M., Castelo-Branco, M., & Januário, C. (2016). *F9 Everyday executive function in huntington's disease: early deficits assessed by a virtual reality task*. BMJ Publishing Group Ltd.

Júlio, F., Ribeiro, M. J., Patrício, M., Malhão, A., Pedrosa, F., Gonçalves, H., Simões, M., van Asselen, M., Simões, M. R., Castelo-Branco, M., & Januário, C. (2019). A novel ecological approach reveals early executive function impairments in Huntington's disease. *Frontiers in Psychology*, *10*(MAR), 1–19. https://doi.org/10.3389/fpsyg.2019.00585

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology Review*, *17*(3), 213–233.

Katz, M. H. (2011). *Multivariable analysis: a practical guide for clinicians and public health researchers*. Cambridge university press.

Klinger, E., Chemin, I., Lebreton, S., & Marie, R. M. (2004). A virtual supermarket to assess cognitive planning. *Annual Review of CyberTherapy and Telemedicine*, *2*, 49–57.

Klinger, E., Chemin, I., Lebreton, S., & Marié, R.-M. (2006). Virtual action planning in parkinson's disease: AControl study. *Cyberpsychology & Behavior*, *9*(3), 342–347.

Krabbe, P. F. M. (2017). Chapter 7 - Validity. In P. F. M. Krabbe (Ed.), *The Measurement of Health and Health Status* (pp. 113–134). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-801504-9.00007-6

Kudlicka, A., Clare, L., & Hindle, J. v. (2011). Executive functions in Parkinson's disease: Systematic review and meta-analysis. *Movement Disorders*, *26*(13), 2305–2315.

la Paglia, F., la Cascia, C., Cipresso, P., Rizzo, R., Francomano, A., Riva, G., & la Barbera, D. (2014). Psychometric assessment using classic neuropsychological and virtual reality based test: A study in obsessive-compulsive disorder (OCD) and schizophrenic patients. *International Symposium on Pervasive Computing Paradigms for Mental Health*, 23–32.

la Paglia, F., la Cascia, C., Rizzo, R., Riva, G., & la Barbera, D. (2016). Decision making and cognitive behavioral flexibility in a ocd sample: a study in a virtual environment. *Stud. Health Technol. Inform*, *219*, 53–57.

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. *Symposium of the Austrian HCI and Usability Engineering Group*, 63–76.

Lawson, R. A., Yarnall, A. J., Duncan, G. W., Breen, D. P., Khoo, T. K., Williams-Gray, C. H., Barker, R. A., Collerton, D., Taylor, J.-P., & Burn, D. J. (2016). Cognitive decline and quality of life in incident Parkinson's disease: the role of attention. *Parkinsonism & Related Disorders*, *27*, 47–53.

Leroi, I., McDonald, K., Pantula, H., & Harbishettar, V. (2012). Cognitive impairment in Parkinson disease: impact on quality of life, disability, and caregiver burden. *Journal of Geriatric Psychiatry and Neurology*, *25*(4), 208–214.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators & Virtual Environments*, *10*(3), 282–297.

Levine, B., Stuss, D. T., Winocur, G., Binns, M. A., Fahy, L., Mandic, M., Bridges, K., & Robertson, I. H. (2007). Cognitive rehabilitation in the elderly: Effects on strategic behavior in relation to goal management. *Journal of the International Neuropsychological Society: JINS*, *13*(1), 143.

Levy, G., Jacobs, D. M., Tang, M., Côté, L. J., Louis, E. D., Alfaro, B., Mejia, H., Stern, Y., & Marder, K. (2002). Memory and executive function impairment predict dementia in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *17*(6), 1221–1226.

Lewis, J. R., & Sauro, J. (2009). The factor structure of the system usability scale. *International Conference on Human Centered Design*, 94–103.

Lezak, M. D. (1995). Executive functions and motor performance. *In Neuropsychological Assessment*, 650–685.

Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). *Neuropsychological assessment*. Oxford University Press, USA.

Luria, A. R. (1973). The frontal lobes and the regulation of behavior. In *Psychophysiology of the frontal lobes* (pp. 3–26). Elsevier.

Luria, A. R., Karpov, B. A., & Yarbuss, A. L. (1966). Disturbances of active visual perception with lesions of the frontal lobes. *Cortex*, *2*(2), 202–212.

Maggio, M. G., de Cola, M. C., Latella, D., Maresca, G., Finocchiaro, C., la Rosa, G., Cimino, V., Sorbera, C., Bramanti, P., & de Luca, R. (2018). What about the role of virtual reality in Parkinson disease's cognitive rehabilitation? Preliminary findings from a randomized clinical trial. *Journal of Geriatric Psychiatry and Neurology*, *31*(6), 312–318.

Maggio, M. G., Russo, M., Cuzzola, M. F., Destro, M., la Rosa, G., Molonia, F., Bramanti, P., Lombardo, G., de Luca, R., & Calabrò, R. S. (2019). Virtual reality in multiple sclerosis rehabilitation: A review on cognitive and motor outcomes. *Journal of Clinical Neuroscience*, *65*, 106–111. https://doi.org/10.1016/j.jocn.2019.03.017

Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain Injury*, *18*(11), 1067–1081.

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316.

McLeod, S. (2018). *Theories of selective attention*.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, *77*(12), 1321–1329.

Milner, B. (1963). Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of Neurology*, *9*(1), 90–100.

Miyake, A., & Shah, P. (1997). *Models of working memory*. COLORADO UNIV AT BOULDER DEPT OF PSYCHOLOGY.

Monaco, M., Costa, A., Caltagirone, C., & Carlesimo, G. A. (2013). Forward and backward span for verbal and visuo-spatial data: standardization and normative data from an Italian adult population. *Neurological Sciences*, *34*(5), 749–754.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & 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.

Nebel, K., Wiese, H., Seyfarth, J., Gizewski, E. R., Stude, P., Diener, H.-C., & Limmroth, V. (2007). Activity of attention related structures in multiple sclerosis patients. *Brain Research*, *1151*, 150–160.

Negro Cousa, E., Brivio, E., Serino, S., Heboyan, V., Riva, G., & 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.

Neguț, A. (2014). Cognitive assessment and rehabilitation in virtual reality : theoretical review and practical implications. *Romanian Journal of Applied Psychology*, *16*(1), 1–7.

Neguț, A., Matu, S.-A., Sava, F. A., & David, D. (2016). Virtual reality measures in neuropsychological assessment: a meta-analytic review. *The Clinical Neuropsychologist*, *30*(2), 165–184.

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, *12*(4), 313–324.

Nir-Hadad, S. Y., Weiss, P. L., Waizman, A., Schwartz, N., & Kizony, R. (2017). A virtual shopping task for the assessment of executive functions: Validity for people with stroke. *Neuropsychological Rehabilitation*, *27*(5), 808–833. https://doi.org/10.1080/09602011.2015.1109523

Norman, D. A., & Shallice, T. (1986). Attention to action. In *Consciousness and self-regulation* (pp. 1–18). Springer.

Novelli, G., Papagno, C., Capitani, E., & Laiacona, M. (1986). Tre test clinici di memoria verbale a lungo termine: taratura su soggetti normali. *Archivio Di Psicologia, Neurologia e Psichiatria*.

Nys, G. M. S., van Zandvoort, M. J. E., de Kort, P. L. M., Jansen, B. P. W., de Haan, E. H. F., & Kappelle, L. J. (2007). Cognitive disorders in acute stroke: prevalence and clinical determinants. *Cerebrovascular Diseases*, *23*(5–6), 408–416.

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), 1–19. https://doi.org/10.3389/fnhum.2015.00660

Parsons, T. D., Courtney, C. G., Arizmendi, B., & Dawson, M. (2011). Virtual reality stroop task for neurocognitive assessment. *Studies in Health Technology and Informatics*, *163*, 433–439. https://doi.org/10.3233/978-1-60750-706-2-433

Parsons, T. D., & Phillips, A. S. (2016). Virtual reality for psychological assessment in clinical practice. *Practice Innovations*, *1*(3), 197–217. https://doi.org/10.1037/pri0000028

Paulwoods, S., & Tröster, A. I. (2003). Prodromal frontal/executive dysfunction predicts incident dementia in Parkinson's disease. *Journal of the International Neuropsychological Society*, *9*(1), 17–24.

Pedroli, E., Cipresso, P., Serino, S., Albani, G., & Riva, G. (2013). A Virtual Reality Test for the assessment of cognitive deficits: Usability and perspectives. *Proceedings of the 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, PervasiveHealth 2013*, 453–458. https://doi.org/10.4108/icst.pervasivehealth.2013.252359

Pedroli, E., Greci, L., Colombo, D., Serino, S., Cipresso, P., Arlati, S., Mondellini, M., Boilini, L., Giussani, V., Goulene, K., Agostoni, M., Sacco, M., Stramba-Badiale, M., Riva, G., & Gaggioli, A. (2018). Characteristics, usability, and users experience of a system combining cognitive and physical therapy in a virtual environment: Positive bike. *Sensors (Switzerland)*, *18*(7). https://doi.org/10.3390/s18072343

Pedroli, E., la Paglia, F., Cipresso, P., la Cascia, C., Riva, G., & la Barbera, D. (2019). A Computational Approach for the Assessment of Executive Functions in Patients with Obsessive–Compulsive Disorder. *Journal of Clinical Medicine*, *8*(11), 1975. https://doi.org/10.3390/jcm8111975

Pedroli, E., Serino, S., Giglioli, A. C., Pallavicini, F., Cipresso, P., & Riva, G. (2016). *The Use of Virtual Reality Tools for the Assessment of Executive Functions and Unilateral Spatial Neglect*. *January*, 115–140. https://doi.org/10.4018/978-1-4666-9740-9.ch007

Pieri, L., Serino, S., Cipresso, P., Mancuso, V., Riva, G., & Pedroli, E. (2021). The ObReco-360°: a new ecological tool to memory assessment using 360° immersive technology. *Virtual Reality*, *0123456789*. https://doi.org/10.1007/s10055-021-00526-1

Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., Obeso, J., Marek, K., Litvan, I., & Lang, A. E. (2015). MDS clinical diagnostic criteria for Parkinson's disease. *Movement Disorders*, *30*(12), 1591–1601.

Rand, D., Rukan, S. B. A., Weiss, P. L., & Katz, N. (2009). Validation of the Virtual MET as an assessment tool for executive functions. *Neuropsychological Rehabilitation*, *19*(4), 583–602. https://doi.org/10.1080/09602010802469074

Raspelli, S., Carelli, L., Morganti, F., Albani, G., Pignatti, R., Mauro, A., Poletti, B., Corra, B., Silani, V., & Riva, G. (2009). A neuro vr-based version of the multiple errands test for the assessment of executive functions: A possible approach. *Journal of Cyber Therapy and Rehabilitation*, *2*(4), 299–314.

Raven, J. C. (1938). *Progressive Matrices: Sets A, B, C, D, and E.* University Press, publiched by HK Lewis.

Raven, J. C. (1947). Progressive matrices: sets A, AbB. *Board and Book Forms*.

Raz, N. (2000). *Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings.*

Realdon, O., Serino, S., Savazzi, F., Rossetto, F., Cipresso, P., Parsons, T. D., Cappellini, G., Mantovani, F., Mendozzi, L., Nemni, R., Riva, G., & Baglio, F. (2019). An ecological measure to screen executive functioning in MS: the Picture Interpretation Test (PIT) 360°. *Scientific Reports*, *9*(1), 1–8. https://doi.org/10.1038/s41598-019-42201-1

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*(3), 271–276.

Reitan, R. M. (1992a). *Trail making test*. Reitan Neuropsychology Laboratory.

Reitan, R. M. (1992b). *Trail Making Test: Manual for administration and scoring*. Reitan Neuropsychology Laboratory.

Renison, B., Ponsford, J., Testa, R., Richardson, B., & Brownfield, K. (2012). The ecological and construct validity of a newly developed measure of executive function: The virtual library task. *Journal of the International Neuropsychological Society*, *18*(3), 440–450. https://doi.org/10.1017/S1355617711001883

Riva, G. (2009). Virtual reality: an experiential tool for clinical psychology. *British Journal of Guidance & Counselling*, *37*(3), 337–345.

Rizzo, A. A., Buckwalter, J. G., Bowerly, T., van der Zaag, C., Humphrey, L., Neumann, U., Chua, C., Kyriakakis, C., van Rooyen, A., & Sisemore, D. (2000). The virtual classroom: A virtual reality environment for the assessment and rehabilitation of attention deficits. *Cyberpsychology and Behavior*, *3*(3), 483–499. https://doi.org/10.1089/10949310050078940

Rizzo, A. A., Buckwalter, J. G., McGee, J. S., Bowerly, T., Zaag, C. van der, Neumann, U., Thiebaux, M., Kim, L., Pair, J., & Chua, C. (2001). Virtual environments for assessing and rehabilitating cognitive/functional performance a review of projects at the USC integrated media systems center. *Presence: Teleoperators & Virtual Environments*, *10*(4), 359–374.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758.

Santangelo, G., Siciliano, M., Pedone, R., Vitale, C., Falco, F., Bisogno, R., Siano, P., Barone, P., Grossi, D., & Santangelo, F. (2015). Normative data for the Montreal Cognitive Assessment in an Italian population sample. *Neurological Sciences*, *36*(4), 585–591.

Sauer, J., Sonderegger, A., & Schmutz, S. (2020). Usability, user experience and accessibility: towards an integrative model. *Ergonomics*, *63*(10), 1207–1220. https://doi.org/10.1080/00140139.2020.1774080

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2014). Applying the user experience questionnaire (UEQ) in different evaluation scenarios. *International Conference of Design, User Experience, and Usability*, 383–392.

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Construction of a Benchmark for the User Experience Questionnaire (UEQ). *Int. J. Interact. Multim. Artif. Intell.*, *4*(4), 40–44.

Schrepp, M., & Thomaschewski, J. (2019). Design and Validation of a Framework for the Creation of User Experience Questionnaires. *International Journal of Interactive Multimedia & Artificial Intelligence*, *5*(7).

Schultheis, M. T., & Rizzo, A. A. (2001). The application of virtual reality technology in rehabilitation. *Rehabilitation Psychology*, *46*(3), 296.

Serino, S., Baglio, F., Rossetto, F., Realdon, O., Cipresso, P., Parsons, T. D., Cappellini, G., Mantovani, F., de Leo, G., Nemni, R., & 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

Serino, S., Baglio, F., Rossetto, F., Realdon, O., Cipresso, P., Parsons, T. D., Cappellini, G., Mantovani, F., de Leo, G., Nemni, R., Riva, G., Burger, G., Guna, J., Pogačnik, M., Lee, N., Choi, W., Lee, S., Sauer, J., Sonderegger, A., … Riva, G. (2020). Usability issues of clinical and research applications of virtual reality in older people: A systematic review. *Sensors (Switzerland)*, *18*(1), 1–10. https://doi.org/10.3389/fnhum.2020.00093

Serino, S., Pedroli, E., Cipresso, P., Pallavicini, F., Albani, G., Mauro, A., & Riva, G. (2014). The role of virtual reality in neuropsychology: The virtual multiple errands test for the assessment of executive functions in Parkinson's disease. *Intelligent Systems Reference Library*, *68*, 257–274. https://doi.org/10.1007/978-3-642-54816-1_14

Serino, S., & Repetto, C. (2018). New trends in episodic memory assessment: immersive 360 ecological videos. *Frontiers in Psychology*, *9*, 1878.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *298*(1089), 199–209.

Shallice, T., & 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

Sirigu, A., Zalla, T., Pillon, B., Grafman, J., Agid, Y., & Dubois, B. (1995). Selective impairments in managerial knowledge following pre-frontal cortex damage. *Cortex*, *31*(2), 301–316.

Spinnler, H., & Tognoni, G. (1987a). Italian Group on the Neuropsychological Study of Ageing: Italian standardization and classification of neuropsychological tests. *Ital J Neurol Sci*, *6*(suppl 8), 1–120.

Spinnler, H., & Tognoni, G. (1987b). Standardizzazione e taratura italiana di tests neuropsicologici.[Italian normative values and standardization of neuropsychological tests]. *Italian Journal of Neurological Sciences*, *6*(Suppl. 8), 1–20.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643.

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: a conceptual view. *Psychological Research*, *63*(3–4), 289–298.

Stuss, D. T., & Benson, D. F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin*, *95*(1), 3.

Tarnanas, I., Schlee, W., Tsolaki, M., Müri, R., Mosimann, U., & Nef, T. (2013). Ecological validity of virtual reality daily living activities screening for early dementia: Longitudinal study. *Journal of Medical Internet Research*, *15*(8), 1–14. https://doi.org/10.2196/games.2778

Tekin, S., & Cummings, J. L. (2002). Frontal–subcortical neuronal circuits and clinical neuropsychiatry: an update. *Journal of Psychosomatic Research*, *53*(2), 647–654.

Tuena, C., Pedroli, E., Trimarchi, P. D., Gallucci, A., Chiappini, M., Goulene, K., Gaggioli, A., Riva, G., Lattanzio, F., Giunco, F., & 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*(April). https://doi.org/10.3389/fnhum.2020.00093

Vaughan, L., & Giovanello, K. (2010). Executive function in daily life: Age-related influences of executive processes on instrumental activities of daily living. *Psychology and Aging*, *25*(2), 343.

Ventura, S., Brivio, E., Riva, G., & Baños, R. M. (2019). Immersive versus non-immersive experience: Exploring the feasibility of memory assessment through 360 technology. *Frontiers in Psychology*, *10*, 2509.

Venturini, R., Radice, M. L., & Imperiali, M. G. (1983). *Il color word test o test di Stroop*. Organizzazioni speciali.

Violante, M. G., Vezzetti, E., & Piazzolla, P. (2019). Interactive virtual technologies in engineering education: Why not 360° videos? *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *13*(2), 729–742.

Werner, P., Rabinowitz, S., Klinger, E., Korczyn, A. D., & Josman, N. (2009). Use of the virtual action planning supermarket for the diagnosis of mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders*, *27*(4), 301–309. https://doi.org/10.1159/000204915

Wilson, B. A., Evans, J. J., Alderman, N., Burgess, P. W., & Emslie, H. (1997). Behavioural assessment of the dysexecutive syndrome. *Methodology of Frontal and Executive Function*, *239*, 250.