Immersive Virtual Reality Methods in Cognitive Neuroscience and Neuropsychology: The Virtual Reality Everyday Assessment Lab (VR-EAL), An Immersive Neuropsychological Test Battery of Everyday Cognitive Functions.

Panagiotis Kourtesis

Thesis presented for the degree of Doctor of Philosophy

University of Edinburgh

2020

Declaration

I, the author and candidate, declare:

(a) that the thesis has been composed by the candidate, and

(b) either that the work is the candidate's own, or, if the candidate has been a member of a research group, that the candidate has made a substantial contribution to the work, such contribution being clearly indicated, and

(c) that the work has not been submitted for any other degree or professional qualification except as specified, and

(d) that any included publications are the candidate's own work, except where indicated throughout the thesis and summarised and clearly identified on the declarations page of the thesis.

Signature: P. Kourtesis Date: 27/7/2020

Table of Contents

Acknowledgments

Above all, I would like to thank the participants of the studies of this thesis, who made this research possible.

I would like to express my sincere gratitude to my mother and my role models (Dr Sarah MacPherson and Professor Sharon Abrahams) who taught me how to be a good and ethical scientist and individual:

Thank you deeply, Dr Sarah MacPherson and Professor Sharon Abrahams for your insights, professionalism, and ethos.

Also, I would like to thank my supervisors, Dr Sarah MacPherson, Professor Simona Collina, and Dr Leonidas Doumas for the indispensable personal and professional guidance, as well as the practical and psychological support.

Lastly, I would like to thank Antonia Felliou, my best friend, partner, and number one supporter, for the continual support and encouragement.

Being an individual who is an immigrant with learning disabilities, I would like to dedicate this thesis to all immigrants who left their home-countries and families pursuing a better future, as well as to all individuals with learning disabilities who work hard to achieve their dreams.

“*Labor Omnia Vincit*”

“*Hard work conquers all*”

Virgil, ancient Roman poet, 70 – 19 B.C.

Acronyms

AACN – American Academy of Clinical Neuropsychology

ABI – Acquired Brain Injuries

AD – Alzheimer’s Disease

ADL – Activities of Daily Living

APA – American Psychological Association

BADS – Behavioral Assessment of Dysexecutive Syndrome

BF – Bayesian Factor

CAMPROMPT – Cambridge Prospective Memory Test

CANTAB – Cambridge Neuropsychological Test Automated Battery

CFA – Confirmatory Factor Analysis

CFI – Confirmatory Factor Index

CNAD – Computerized Neuropsychological Assessment Devices

CTT – Color Trails Test

CV – Commercial Version

DK – Development Kit

DoF – Degrees of Freedom

EVET – Edinburgh Virtual Errands Test

FOV – Field of View

GE – Game Engine

HIV – Human Immunodeficiency Viruses

HMD – Head-Mounted Display

IADL – Instrumental Activities of Daily Living

JZS – Jeffrey–Zellner–Siow

LCD – Liquid Crystal Display

MET – Multiple Errands Test

MoCAP – Motion Capture

MP – Multiprocess theory

NAN – National Academy of Neuropsychology

NPC – Non-Player Characters

OLED – Organic Light Emitting Diode

PAM – Preparatory Attentional and Memory processes

PD – Parkinson’s Disease

PM – Prospective Memory

RBMT-III – Rivermead Behavioral Memory Test–III

RMSEA – Root Mean Squared Error of Approximation

RSAT – Ruff Selective Attention Test

SDK – Software Development Kit

SPSS – Statistical Package for the Social Sciences

SRMR – Standardized Root Mean square Residual

SSQ – Simulator Sickness Questionnaire

TBI – Traumatic Brain Injuries

TEA – Test of Everyday Attention

TMT – Trail Making Test

TLI – Tuckere Lewis Index

VE – Virtual Environment

VM – Virtual Mall

VR – Virtual Reality

VR-EAL – Virtual Reality Everyday Assessment Lab

VRISE – Virtual Reality Induced Symptoms and Effects

VMET – Virtual Multiple Errands Test

VRNQ – Virtual Reality Neuroscience Questionnaire

VRSQ – Virtual Reality Sickness Questionnaire

VRTK – Virtual Reality Toolkit

WCST – Wisconsin Card Sorting Test

WMS-R – Wechsler Memory Scale-Revised

Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Abstract

In cognitive neuroscience and neuropsychology, the collection of cognitive and behavioural data is predominantly achieved by implementing paper-and-pencil and computerized (i.e., 2D and 3D applications) assessments. However, these psychometric tools in clinics and/or laboratories display several limitations and discrepancies between the observed performance in the laboratory/clinic and the actual performance of individuals in everyday life. The functional and predictive association between an individual's performance on a set of neuropsychological tests and the individual's performance in various everyday life settings is called ecological validity. Ecological validity is considered an important issue that cannot be resolved by the currently available assessment tools. Virtual reality head-mounted displays (HMD) appear to be effective research tools, which may address the problem of ecological validity in neuropsychological testing. However, their widespread implementation is hindered by virtual reality induced symptoms and effects (VRISE) and the lack of skills in virtual reality software development.

In this PhD, a technological systematic literature review of the reasons for adverse symptomatology was conducted and suggestions and technological knowledge for the implementation of virtual reality HMD systems in cognitive neuroscience provided. The review indicated features pertinent to display, sound, motion tracking, navigation, ergonomic interactions, user experience, and computer hardware that should be considered by researchers. Subsequently, a meta-analysis of 44 neuroscientific or neuropsychological studies involving virtual reality HMD systems was performed. The meta-analysis of the virtual reality studies demonstrated that new generation HMDs induce significantly less VRISE and marginally fewer dropouts. Importantly,

the commercial versions of the new generation HMDs with ergonomic interactions had zero incidents of adverse symptomatology and dropouts. HMDs equivalent to or greater than the commercial versions of contemporary HMDs accompanied with ergonomic interactions are suitable for implementation in cognitive neuroscience.

Another aim of this PhD was to devise a brief tool to appraise and report both the quality of software features and VRISE intensity quantitatively; such a tool does not currently exist. The Virtual Reality Neuroscience Questionnaire (VRNQ; Kourtesis et al., 2019) was developed to assess the quality of virtual reality software in terms of user experience, game mechanics, in-game assistance, and VRISE. Forty participants aged between 28 and 43 years were recruited (18 gamers and 22 non-gamers) for the study. They participated in 3 different virtual reality sessions until they felt weary or discomfort and subsequently filled in the VRNQ. The results demonstrated that VRNQ is a valid tool for assessing virtual reality software as it has good convergent, discriminant, and construct validity. The maximum duration of virtual reality sessions should be between 55 and 70 min when the virtual reality software meets or exceeds the parsimonious cut-offs of the VRNQ, and the users are familiarized with the virtual reality system. Also, gaming experience does not affect how long virtual reality sessions should last. Furthermore, while the quality of virtual reality software substantially modulates the maximum duration of virtual reality sessions, age and education do not. Finally, deeper immersion, better quality of graphics and sound, and more helpful in-game instructions and prompts were found to reduce VRISE intensity. The VRNQ facilitates the brief assessment and reporting of the quality of virtual reality software features and/or the intensity of VRISE, while its minimum and parsimonious cut-offs may appraise the suitability of virtual reality software for implementation in research and clinical settings. However, the development of virtual reality software is predominantly dependent on third parties (e.g., freelancers or companies) with programming and software development skills. A solution that will promote the adoption of immersive virtual reality as a research and clinical tool might be the in-house development of virtual reality research/clinical software by computer science literate cognitive scientists or research software engineers.

In Chapter 4, guidelines are offered for the development of virtual reality software in cognitive neuroscience and neuropsychology, by describing and discussing the stages of the development of Virtual Reality Everyday Assessment Lab (VR-EAL), the first neuropsychological battery in immersive virtual reality. Techniques for evaluating cognitive functions within a realistic storyline are discussed. The utility of various assets in Unity, software development kits, and other software are described so that cognitive scientists can overcome challenges pertinent to VRISE and the quality of the virtual reality software. In addition, VR-EAL is evaluated in accordance with the necessary criteria for virtual reality software for research purposes. The virtual reality neuroscience questionnaire (VRNQ) was implemented to appraise the quality of the three versions of VR-EAL in terms of user experience, game mechanics, in-game assistance, and VRISE. Twenty-five participants aged between 20 and 45 years with 12–16 years of full-time education evaluated various versions of VR-EAL. The final version of VR-EAL achieved high scores in every sub-score of the VRNQ and exceeded its parsimonious cut-offs. It also appeared to have better in-game assistance and game mechanics, while its improved graphics substantially increased the quality of the user experience and almost eradicated VRISE. The results substantially support

the feasibility of the development of effective virtual reality research and clinical software without the presence of VRISE during a 60-min virtual reality session.

In Chapter 5, validation of VR-EAL as an assessment of prospective memory, episodic memory, attention, and executive functions using an ecologically valid approach is examined. Performance on the VR-EAL, an immersive virtual reality neuropsychological battery, is examined against an extensive paper-and-pencil neuropsychological battery. Forty-one participants (21 females) were recruited: 18 gamers and 23 non-gamers who attended both an immersive virtual reality and a paper-and-pencil testing session. Bayesian Pearson correlation analyses were conducted to assess construct and convergent validity of the VR-EAL. Bayesian t-tests were performed to compare virtual reality and paper-and-pencil testing in terms of administration time, similarity to real life tasks (i.e., ecological validity), and pleasantness. VR-EAL scores were significantly correlated with their equivalent scores on the paper-and-pencil tests. The participants' reports indicated that the VR-EAL tasks were considered significantly more ecologically valid and pleasant than the paper-and-pencil neuropsychological battery. The VR-EAL battery also had a shorter administration time. The VR-EAL appears to be an effective neuropsychological tool for the assessment of everyday cognitive functions, and has enhanced ecological validity, a highly pleasant testing experience, and does not induce cybersickness.

In the final part of this thesis, the preparatory attentional and memory (PAM) and the multiprocess theories of prospective memory are examined by attempting to identify the cognitive functions which may predict the individual's performance on ecologically valid prospective memory tasks in the same group of participants described in Chapter 5. Bayesian t-tests were conducted to explore the differences

among different prospective memory tasks (e.g., event-based and time-based) and prospective memory tasks with varying delays between encoding and the recall of the intended action (e.g., short-delay versus long-delay). Bayesian linear regression analyses were performed to examine the predictors of VR-EAL scores. The results revealed that the type of prospective memory task does not play a significant role in everyday prospective memory functioning, but instead the length of delay between encoding and retrieving the prospective memory intention plays a central role. Support for the PAM and MP frameworks was found in non-focal and focal event-based tasks respectively. However, the findings, inferring a dynamic interplay between automatic and intentional monitoring and retrieval processes, agree with the inclusive approach of the multiprocess framework. Also, the role of executive functions appears crucial in everyday PM. Finally, everyday PM is predominantly facilitated by episodic memory, visuospatial attention, auditory attention, and executive functions.

In conclusion, this PhD thesis attempted to show how immersive virtual reality research methods may be implemented efficiently without the confounding effect of cybersickness symptomatology in order to enhance the ecological validity of neuropsychological testing and contribute to our understanding of everyday cognitive ability.

The following chapters are published in international journals:

Chapter 2: Kourtesis, P., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2019). Technological competence is a precondition for effective implementation of virtual reality head mounted displays in human neuroscience: a technological review and meta-analysis. *Frontiers in Human Neuroscience*, 13, 342.

Chapter 3: Kourtesis, P., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2019). Validation of the virtual reality neuroscience questionnaire: maximum duration of immersive virtual reality sessions without the presence of pertinent adverse symptomatology. *Frontiers in Human Neuroscience*, 13, 417.

Chapter 4: Kourtesis, P., Korre, D., Collina, S., Doumas, L. A., & MacPherson, S. E. (2020). Guidelines for the development of immersive virtual reality software for cognitive neuroscience and neuropsychology: the development of virtual reality everyday assessment lab (VR-EAL), a neuropsychological test battery in immersive virtual reality. *Frontiers in Computer Science*, 1, 12.

Chapter 5 is accepted for publication: Kourtesis, P., Collina, S., Doumas, L. A., & MacPherson, S. E. (2020). Validation of the Virtual Reality Everyday Assessment Lab (VR-EAL): An immersive virtual reality neuropsychological battery with enhanced ecological validity. *Journal of the International Neuropsychological Society*, in press.

Chapter 6 is submitted for publication: Kourtesis, P., Collina, S., Doumas, L. A., & MacPherson, S. E. (submitted). An ecologically valid examination of the theoretical framework of prospective memory: the facilitation of real-world prospective memory by attention, memory, and executive function processes. *Memory & Cognition*.

Lay Summary

Everyday functioning relies on various cognitive abilities such as memory, attention, planning, and multitasking. One of the most essential cognitive abilities in everyday life is the ability to remember the intention to perform a planned action in the future. To depict the quality of an individual's everyday functioning, the evaluation of these cognitive abilities should be performed in a way where the testing procedure resembles the complexity and demands of everyday situations. While there are several tests of these cognitive abilities in paper-and-pencil or digital form, they suffer from certain limitations which prevent them from accurately portraying the everyday functioning of individuals. Contemporary virtual reality systems appear capable of overcoming these limitations and provide tests which adequately resemble everyday situations.

However, there are some pitfalls in utilizing virtual reality systems for research and clinical purposes. The main problem is that virtual reality systems may induce adverse symptoms such as nausea, dizziness, disorientation, instability, and fatigue. This symptomatology is frequently referred as virtual reality induced symptoms and effects (VRISE). This thesis aimed to provide methods for the avoidance or mitigation of VRISE and develop a virtual reality test of everyday cognitive abilities under the name Virtual Reality Everyday Assessment Lab (VR-EAL) capable of contributing to our understanding of everyday cognitive functioning.

Through a series of studies, the above-mentioned aims were achieved. The technological reasons of VRISE, pertaining to hardware and software features, were identified. The identification of the required hardware characteristics assisted with providing guidelines for the selection of appropriate virtual reality headsets which do not induce VRISE. Similarly, the identification of software features able to mitigate VRISE intensity and frequency assisted with the development of VR-EAL. However, there was not any available tool (i.e., questionnaire) which appraises the quality of these software features and VRISE intensity. For this reason, the Virtual Reality Neuroscience Questionnaire (VRNQ) was developed and validated as a tool which measures the intensity of VRISE, as well as the quality of software features known for mitigating VRISE.

Furthermore, this thesis, based on the VR-EAL's development, provided guidelines for the development of virtual reality software for research and clinical purposes. Using the VRNQ, the VR-EAL was found to not inducing VRISE in 66 individuals which participated in two studies (i.e., 25 and 41 in each study). The VR-EAL was also rated by 41 participants as substantially more pleasant testing experience and more closely resembling everyday situations than infamous paper-and-pencil tests of the same cognitive abilities. Finally, the VR-EAL was able to identify cognitive abilities which are required for remembering the intention to perform a planned action in the future. Consequently, the VR-EAL appears able to contribute to our understanding of everyday cognitive functioning.

Chapter 1: Introduction

Everyday functioning relies on several cognitive functions, which include prospective memory, episodic memory, attentional processes, and executive functions. The neuropsychological assessment of these cognitive functions is thought to benefit from ecological validity. This chapter provides an overview of the relationship of these central cognitive functions with everyday functioning, the approaches adopted to achieve ecological validity (i.e., verisimilitude, veridicality), existing ecologically valid tests (i.e., paper-and-pencil, real-world, and digital) and their limitations, and the potentials of immersive virtual reality tests to overcome these challenges and provide enhanced ecological validity.

1.1. Everyday cognition and functioning

Individuals perform various activities in their everyday life which allow them to be independent and functional members of their society. These activities may be majorly classified into two categories: the basic activities of daily living (basic ADL) and the instrumental activities of daily living (IADL; Mlinac & Feng, 2016). The basic ADL refer to the management of essential physical needs such as personal hygiene, dressing, toileting, walking/moving, and eating (Mlinac & Feng, 2016). On the other hand, the IADL refer to more complex everyday activities such as the management of finances and medication, commuting, driving, shopping, working, and socializing (Mlinac & Feng, 2016). Impaired basic ADL are predominantly due to impaired physical functioning such as motor disabilities, internal organ failures, and physiological dysfunction (Boyle, Cohen, Paul, Moser, & Gordon, 2002; Cahn-Weiner *et al.*, 2007). In contrast, impaired IADL are chiefly due to cognitive impairment and decline (Boyle et al, 2002; Cahn-Weiner *et al.*, 2007). In clinical conditions affecting cognition, IADL are usually impaired in the early stages of the condition, while basic ADL become impaired in the later stages of the condition (Mlinac & Feng, 2016). For example, impaired IADL (e.g., managing household finances) are present in mild cognitive impairment or the early stages of dementia (Farias *et al.*, 2013), while impaired basic ADL (e.g., maintaining personal hygiene) emerge in the later stages of dementia (Cahn-Weiner *et al.*, 2007; West, McCue, & Golden, 2012). The cognitive ability of the individual hence appears to be intrinsically related to both basic ADL and IADL abilities, although the former is frequently impaired due to physical dysfunction.

Everyday functioning is thought to be facilitated by cognitive abilities such as attention, episodic memory, prospective memory, and executive functions (Chaytor &

Schmitter-Edgecombe, 2003; Haines *et al.*, 2019; Higginson, Arnett, & Voss, 2000; Mlinac & Feng, 2016; Phillips, Henry, & Martin, 2012; Rosenberg, 2015). Attentional processes have been found to be strongly associated with both ADL and IADL (Freilich & Hyer, 2007; Hall, Vo, Johnson, Barber, & O'Bryant, 2011; Higginson *et al.*, 2000; Ruff *et al.*, 1993). Similarly, immediate and delayed episodic memory have been found to predict basic ADL and IADL (Freilich & Hyer, 2007; Goldstein, McCue, Rogers, & Nussbaum, 1992; Jefferson *et al.*, 2006; Farias, Harrell, Neumann, & Houtz, 2002). Executive functions such as inhibition, task-shifting, and planning have also been associated with and predict basic ADL and IADL (Hall *et al.*, 2011; Johnson, Lui, & Yaffe, 2007; Martyr & Clare, 2012), as well as everyday cognitive performance (e.g., everyday memory tasks such as remembering to take medication; Norris & Tate, 2000). Furthermore, visuospatial perceptual skills showed a strong relationship with basic ADL functioning (Freilich & Hyer, 2007; Perry & Hodges, 2000; Warrington & James, 1991). Lastly, prospective memory was found to be an important predictor of IADL, especially with activities related to occupational performance (Honan, Brown, & Batchelor, 2015).

Notably, prospective memory appears to be particularly crucial in IADL (Crovitz & Daniel, 1984; Einstein & McDaniel, 1996; Kidder, Park, Hertzog, & Morrell, 1997). More than half of everyday memory failures were found to be related to prospective memory functioning (e.g., forgetting to take medication on time; Crovitz & Daniel, 1984; Einstein & McDaniel, 1996). Prospective memory has been found to be one or the only predictor of IADL in several patient groups including patients with mild cognitive impairment (Schmitter-Edgecombe, Woo, & Greeley, 2009), patients with brain injuries (Groot, Wilson, Evans, & Watson, 2002), human immunodeficiency virus (HIV; Woods *et al.*, 2008), schizophrenia (Twamley *et al.*, 2008), and Parkinson's disease (Pirogovsky, Woods, Filoteo, & Gilbert, 2012).

While prospective memory functioning clearly plays a role in IADL, it has been found to be dependent on a number of cognitive functions such as episodic memory (Einstein & McDaniel, 1996; Mackinlay, Kliegel, & Mäntylä, 2009; McFarland & Glisky, 2009), visual attention (Smith, 2003; Smith, Hunt, McVay, & McConnell, 2007), auditory attention (McDaniel & Scullin, 2010), and executive functions (Azzopardi, Auffray, & Kermarrec, 2017; Gonneaud *et al.*, 2011; Schnitzspahn, Stahl, Zeintl, Kaller, & Kliegel, 2013; Zuber, Kliegel, & Ihle, 2016; Zuber, Mahy, & Kliegel, 2019). As discussed above, the cognitive functions which facilitate prospective memory are also facilitators of everyday functioning. Consequently, the assessment of prospective memory functioning and these associated cognitive functions could also inform individuals' everyday functioning.

1.2. Prospective memory and everyday functioning

Prospective memory refers to the ability to remember to perform a particular action in the future (Brandimonte, Einstein, & McDaniel, 2014). Prospective memory can be related to a particular event (e.g., when you are on your lunch break, call your family doctor to ask about your blood test results) or a specific time (e.g., at 4 pm you need to pick up your child from after-school club; Einstein & McDaniel, 1996). These are referred as event-based and time-based prospective memory respectively (Einstein & McDaniel, 1996). Prospective memory is explained by two main theoretical frameworks, the preparatory attentional and memory processes framework (PAM) and the multiprocess framework (Anderson, McDaniel, & Einstein, 2017). The PAM framework postulates that prospective memory functioning requires a constant top-down monitoring for environmental and internal cues which facilitates the retrieval of the intention to perform the prospective memory action (Smith, 2003; Smith *et al.*, 2007). For example, if an individual wants to buy a pint of milk before going back home after work, they will be vigilant and monitoring the environment for cues (e.g., the sign of a supermarket) that will remind them of the intention to buy a pint of milk. The multiprocess framework offers a complementary view to the PAM’s notion of strategic monitoring and retrieval (McDaniel & Einstein, 2000, 2007). The multiprocess framework argues that prospective memory functioning is not always a top-down process (i.e., involving strategic monitoring and intentional retrieval) but also a bottom-up process (i.e., involving automatic monitoring and reflexive associative retrieval). In the previous example, the individual may be in a passive state (i.e., not being vigilant). However, on the way home, they see a huge billboard which displays dairy products. This event triggers the reflexive associative retrieval of the intention to buy a pint of milk. Therefore, the main difference between the PAM and multiprocess frameworks pertains to the retrieval of the intention to perform a prospective memory action (e.g., buying a pint of milk). The PAM theory suggests that there is always an intentional retrieval process which is facilitated by strategically monitoring for cues associated with the prospective memory action. In addition to this retrieval process, the multiprocess framework supports the existence of a reflexive associative retrieval process, which is facilitated by passively detecting environmental and internal cues strongly associated with the intended prospective memory action (McDaniel, Umanath, Einstein, & Waldum, 2015).

Differences between the PAM and the multiprocess frameworks also pertain to the focality and salience of the internal or environmental cue which stimulates the retrieval of the prospective memory intention. For example, the intentional monitoring and retrieval proposed by the PAM are both performed when the cue is non-focal (e.g., a convenience store sign among several signs for nearby shops and buildings) and non-salient in relation to the prospective memory task (e.g., a convenience store sign when the intention is to buy a pint of milk; Einstein, McDaniel, Manzi, Cochran, Baker, 2000; Einstein, Smith, McDaniel, Shaw, 1997; McDaniel *et al.*, 2015; Smith, 2003). On the other hand, automatic monitoring and retrieval, which the multiprocess framework adds, are performed when the cue is focal (e.g., an easily detectable advertisement on the side of the road) and salient in relation to the prospective memory task (e.g., the advertisement displays dairy products when the intention is to buy a pint of milk; Einstein *et al.*, 1997, 2000; McDaniel *et al.*, 2015; Smith, 2003). Also, performance on focal prospective memory tasks (i.e., where the cue is focal and salient) has been observed to be significantly better than performance on non-focal prospective memory tasks (i.e., where the cue is non-focal and non-salient; Anderson *et al.*, 2017; McDaniel *et al.*, 2015; Mullet *et al.*. 2013; Scullin, McDaniel, Shelton, & Lee, 2010). Finally, in addition to the significance of attentional processes for prospective memory functioning, the PAM theory also underlines the importance of preparatory processes related to episodic memory (e.g., encoding and maintaining the prospective memory intention) in prospective memory functioning (Smith, 2003; Smith *et al.*, 2007). In contrast, the multiprocess framework focuses on the salience between the cue and prospective memory intention (e.g., falling under the same semantic category such as dairy products and pint of milk) which facilitates more robust encoding and reflexive associative retrieval of the prospective memory intention (Anderson *et al.*, 2017; McDaniel & Einstein, 2000, 2007; McDaniel *et al.*, 2015).

Prospective memory functioning encompasses both event-based and time-based prospective memory tasks, where time-based tasks appear to be more cognitively demanding than event-based tasks (Einstein & McDaniel, 1996). However, the PAM and multiprocess frameworks mainly apply to event-based prospective memory functioning, while they do not explain time-based prospective memory functioning. Also, these frameworks do not explain the role of executive functions in prospective memory functioning, while studies provided evidence postulating a strong relationship between prospective memory and executive functions (e.g., Azzopardi *et al.*, 2017; Gonneaud *et al.*, 2011; Schnitzspahn, *et al.*, 2013; Zuber *et al.*, 2016, 2019). In particular, the executive functions that have been found to contribute to prospective memory are those related to the tripartite model by Miyake and collaborators (2000). This model includes three diverse executive functioning processes, which are updating (i.e., the management of stored information in working memory), inhibition (i.e., the avoidance of responding prematurely and resisting distractions), and shifting (i.e., shifting attention between two or more different tasks; Miyake *et al.*, 2000). These executive functioning processes have been found to contribute to both event-based and time-based PM functioning (Azzopardi *et al.*, 2017; Gonneaud *et al.*, 2011; Schnitzspahn *et al.*, 2013; Zuber *et al.*, 2016; Zuber *et al.*, 2019). Hence, the study of prospective memory should include all types of prospective memory tasks (i.e., focal event-based, non-focal event-based, and time-based) and the executive functioning processes of the tripartite model.

1.3. Neuropsychological assessment of everyday cognition

In cognitive neuroscience and neuropsychology, the acquisition of data pertaining to cognitive functionality is mainly facilitated by the implementation of psychometric tools such as neuropsychological tests and test batteries. The psychometric tools are predominantly restricted to paper-and-pencil and computerized (i.e., 2D and 3D applications) forms. However, neuropsychological assessment benefits from being ecologically valid so that outcomes generalize to individuals' everyday cognitive functioning (Chaytor & Schmitter-Edgecombe, 2003; Haines *et al.*, 2019; Higginson *et al.*, 2000; Mlinac & Feng, 2016; Parsons, 2015; Phillips *et al.*, 2008; Rand, Rukan, Weiss, & Katz, 2009; Rosenberg, 2015). This is because ecologically valid neuropsychological tasks approach the complexity and cognitive demands of equivalent everyday tasks (Franzen & Wilhelm, 1996; Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). Hence, ecological validity augments the probability that the observed cognitive performance of an individual will depict how the individual will perform in real-life conditions (Bailey, Henry, Rendell, Phillips, & Kliegel, 2010; Burgess *et al.*, 2006; Chaytor & Schmitter-Edgecombe, 2003).

As discussed above, the assessment of prospective memory and its associated cognitive functions is thought to inform clinicians and researchers about individuals' cognitive functioning in everyday life. However, the studies examining the relationship between prospective memory and its associated cognitive functions have involved non-ecologically valid tasks (e.g., Einstein *et al.*, 1997, 2000; Mullet *et al.*, 2013; Scullin *et al.*, 2010; Smith, 2003; Smith *et al.*, 2007). These findings cannot be generalized to everyday prospective memory functioning (Chaytor & Schmitter-

Edgecombe, 2003; Haines *et al.*, 2019; Higginson *et al.*, 2000; Mlinac & Feng, 2016; Parsons, 2015; Phillips *et al.*, 2008; Rand *et al.*, 2009; Rosenberg, 2015) and they may provide results which are discrepant from the findings of ecologically valid assessments (Marsh, Hicks, & Landau, 1998). On the contrary, the ecologically valid assessment of prospective memory and its associated cognitive functions would allow an examination of the aspects facilitating everyday prospective memory functioning such as attentional switching, sustained attention, retrospective memory, and metamemory (e.g., strategies to encode and consolidate prospective memory intentions; Marsh *et al.*, 1998). The utilization of ecological valid assessments hence is required for the thorough examination of the complex structure of everyday prospective memory functioning.

Ecological validity can be achieved by adopting one of two approaches: verisimilitude and veridicality (Franzen & Wilhelm, 1996; Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). Verisimilitude pertains to the level that a neuropsychological test resembles the complexity and cognitive demands of a corresponding everyday task (Franzen & Wilhelm, 1996; Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). Following the verisimilitude approach, the development of a neuropsychological test focuses on the creation of tasks, which simulate the procedures and settings of everyday life (Franzen & Wilhelm, 1996; Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). These tests prioritise the identification of individuals who are impaired in performing real-world tasks, rather than assisting in the diagnosis of a clinical condition (e.g., brain damage; Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). Therefore, neuropsychological tests with verisimilitude would be expected to effectively monitor the changes (i.e., increase or decrease) in everyday functioning (Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006).

Another approach to achieve ecological validity is veridicality. Veridicality pertains to the strength of the statistical relationship between the outcomes of neuropsychological tests and everyday functioning measures (e.g., questionnaires of everyday functioning and independence, occupational or academic performance, and other IADL; Franzen & Wilhelm, 1996; Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). The principal advantage of the veridicality approach is that it may be applied to the existing traditional paper-and-pencil tests, with which the researchers and clinicians are already accustomed (Chaytor & Schmitter-Edgecombe, 2003). The traditional paper-and-pencil tests are effective in assisting with the diagnosis of clinical conditions (e.g., dementia) by assessing certain cognitive processes (Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). However, the existing traditional paper-and-pencil tests do not necessarily predict everyday cognitive performance because their only purpose is to measure a specific cognitive process and isolate any confounding factors (Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006).

As discussed above, using neuropsychological assessments with both verisimilitude and veridicality has its advantages. Indeed, the paper-and-pencil tests with verisimilitude appear to have better face validity (i.e., the degree to which a test appears to measure the cognitive process that it is supposed to measure) than equivalent conventional paper-and-pencil tests (Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). Also, the outcomes of paper-and-pencil tests with verisimilitude appear to explain a greater percentage of the variance of everyday functioning than traditional paper-and-pencil tests, which indicates the superiority of tests with verisimilitude also in terms of veridicality (Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). Finally, neuropsychological tests with verisimilitude may better predict real-world memory and attention (Higginson *et al.*, 2000), executive functioning (e.g., multi-tasking, planning and mental flexibility; Burgess, Alderman, Evans, Emslie, & Wilson, 1998) and prospective memory abilities (i.e., remembering to perform an intended action in the future; Haines *et al.*, 2019; Phillips *et al.*, 2012) than the traditional paper-and-pencil tests with veridicality (Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006).

1.4. Ecologically valid neuropsychological tests

1.4.1. Veridicality and traditional paper-and-pencil tests

As mentioned above, the veridicality of neuropsychological tests is established by examining the statistical relationship between performance on a neuropsychological test and everyday functioning measures such as occupational status, IADL or ADL questionnaires, or clinician reports (Chaytor & Schmitter-Edgecombe, 2003). However, the traditional neuropsychological tests, which have shown veridicality, were not typically developed in pursuit of ecological validity, but instead their design was to accurately examine specific cognitive processes (e.g., cognitive flexibility; Chaytor & Schmitter-Edgecombe, 2003). One of the most commonly used neuropsychological assessments of memory is the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987). The WMS-R was found to be able to classify patients with traumatic brain injuries (TBI) with 50% accuracy (Makatura, Lam, Leahy, Castillo, & Kalpakjian, 1999). Also, the WMS-R explained 10% of the variance in TBI patients' everyday cognitive functioning (Bowman, 1996), and correlated with the everyday memory ability of Alzheimer's patients as reported by their family members (Johnson, 1994). In terms of assessing executive abilities, both the Wisconsin Card Sorting Test (WCST; Berg, 1948) and the Trail Making Test (TMT; Reitan, & Wolfson, 1993) are commonly used. The WCST was found to be significantly correlated with the social interactive skills and planning abilities of schizophrenia patients as rated by their clinicians (Poole, Ober, Shenaut, & Vinogradov, 1999). Similarly, the TMT Part B was a found to be a significant predictor of the everyday executive skills of neurological (Burgess *et al.*, 1998) and TBI patients (Chaytor, Schmitter-Edgecombe, & Burr, 2006). Finally, the Ruff Selective Attention Test was found to be a significant predictor of TBI patients' ability to return to professional or academic environments after rehabilitation (Ruff *et al.*, 1993).

However, these traditional paper-and-pencil tests were not designed with ecological validity in mind (Chaytor & Schmitter-Edgecombe, 2003; Higginson *et al.*, 2000; Spooner & Pachana, 2006). Consequently, they do not provide insightful information pertinent to everyday functionality, since they exclude the effect of the environment (e.g., a supermarket) and other confounding factors (e.g., traffic, crowds, use of smartphones, temperature, and building's lighting; Franzen & Wilhelm, 1996; Heinrichs, 1990; Wilson, 1993). Also, due to the exclusion of confounding factors that may be disruptive in real-life settings, the performance on these tests may be seen improved compared to the real-world functionality of an individual (Franzen & Wilhelm, 1996; Heinrichs, 1990; Wilson, 1993). As a result, the significance of the statistical relationships between traditional paper-and-pencil tests and measures of everyday functionality (e.g., ADL and IADL questionnaires) are inconsistent across the studies examining the ecological validity of neuropsychological tests. Non-significant relationships are reported in some studies and significant relationships in others (Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). Lastly, the variance in the everyday functionality explained by traditional paper-and-pencil tests is substantially smaller than the variance in the everyday functionality explained by paper-and-pencil tests with verisimilitude, indicating that the latter may also have better veridicality (Chaytor & Schmitter-Edgecombe, 2003; Higginson *et al.*, 2000; Spooner & Pachana, 2006).

1.4.2. Computerized tasks

In cognitive neuroscience, laboratory tasks frequently take a computerized form (Gur *et al.*, 2010; Mathôt, Schreij, & Theeuwes, 2012; Peirce, 2007; 2009). The implementation of computerized tasks allows neuropsychological assessment to take place in conjunction with neuroimaging techniques such as functional magnetic resonance imaging and eye-tracking (Gur *et al.*, 2010; Mathôt *et al.*, 2012; Peirce, 2007; 2009). The computerized tasks are developed using open source tools such as PsychoPy (Peirce, 2007; 2009) and OpenSesame (Mathôt *et al.*, 2012). They are often computerized versions of the traditional paper-and-pencil tasks or closely follow the rationale of the corresponding traditional paper-and-pencil tasks (Gur *et al.*, 2010). For example, the ACEmobile is a computerised version of the Addenbrooke's Cognitive Examination-III for mobile devices such as smartphones and tablets (Newman *et al.*, 2018). Another example is the Cambridge Neuropsychological Test Automated Battery (CANTAB), which assesses visual memory, attention, working memory and planning; and it includes tasks which are computerised versions of neuropsychological tests such as the delayed matching-to-sample test, the WCST, and the Tower of London (Sahakian & Owen, 1992).However, the computerized tasks may also be developed to examine specific neurocognitive domains (i.e., specific cognitive processes which may activate certain brain structures and/or regions) because they may isolate confounding variables and permit the assessment of a particular cognitive process (Gur *et al.*, 2010). For example, a computerized laboratory paradigm for studying prospective memory functioning may involve an on-going lexical decision task where the participant is required to respond whether the presented word is an actual or fictional word, and a comparable prospective memory task, where the participant should indicate when a specific word appears (Anderson *et al.*, 2017).

It is hypothesized that the computerized tasks display advantages comparable to the equivalent traditional paper-and-pencil tasks (Arrieux, Cole, & Ahrens, 2017; Cole, Arrieux, Ivins, Schwab, & Qashu, 2018). However, the veridicality of computerised tasks has not yet been explored. Also, the computerised versions may suffer from limitations comparable to the traditional paper-and-pencil tests (e.g., a two-dimensional interface, the exclusion of confounding factors, and facilitating improved performance on these tests), since the computerised versions were not developed to address these issues (Arrieux *et al.*, 2017; Cole *et al.,* 2018). As discussed above, the traditional paper-and-pencil tests display inconsistent relationships with real-world performance, and when a relationship is established, they explain a relatively small amount of variance in everyday functionality (Chaytor & Schmitter-Edgecombe, 2003; Higginson *et al.*, 2000; Spooner & Pachana, 2006)

1.4.3. Ecological valid paper-and-pencil tests with verisimilitude

In an attempt to increase the ecologically validity of neuropsychological assessments using a verisimilitude approach, several paper-and-pencil tests were developed which endeavour to resemble the cognitive demands and complexity of everyday tasks

(Chaytor & Schmitter-Edgecombe, 2003). The most prominent tests are the Test of Everyday Attention (TEA; Robertson, Ward, Ridgeway, & Nimmo-Smith, 1996), the Rivermead Behavioral Memory Test–III (RBMT–III; Wilson, Cockburn, & Baddeley, 2008), the Behavioral Assessment of Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996), and the Cambridge Prospective Memory Test (CAMPROMPT; Wilson, 2005). The TEA measures everyday attention pertaining to visual and auditory attentional processes, as well as attentional control (e.g., shifting attention between two tasks; Robertson *et al.*, 1996). The RBMT-III assesses aspects of everyday episodic memory such as visual and verbal memory, delayed recognition, and immediate and delayed recall (Wilson *et al.*, 2008). The BADS appraises everyday executive functions such as temporal judgement, shifting, inhibition, problem solving, and strategy formation (Wilson *et al.*, 1996). Finally, the CAMPROMPT measures both event-based and time-based prospective memory (Wilson, 2005).

These tests attempt to provide ecological validity (i.e., verisimilitude) by using the testing environment (i.e., the room that the test is being taken) and various props (e.g., a map of a city, and a vase filled with water). Both CAMPROMPT and RBMT-III use the testing environment; for example, they both include a task where the examinee should remember the location of objects (e.g., a pen, pencil, or rubber) that were hidden in various places in the room (e.g., on the corner, under the desk, or in the bin; Wilson *et al.*, 1996, 2008). Furthermore, examples of using props to achieve verisimilitude can be seen in the TEA and BADS. For example, the TEA requires the examinee to scan a map of Philadelphia, USA and detect all the symbols which indicate a restaurant or a gas station (Robertson *et al.*, 1996), while the BADS requires the participant to remove a cork from a vase filled with water, only by using the available tools (e.g., a wire) and without touching the vase (Wilson *et al.*, 1996).

Importantly, these paper-and-pencil tests with verisimilitude have also been found to be ecologically valid in terms of veridicality, which supports their ability to predict everyday functioning. For example, the RBMT was able to predict the everyday memory functioning of TBI patients (Makatura *et al.*, 1999). Also, the RBMT showed a strong correlation with occupational therapists' observations of ADL (e.g., personal hygiene and medication) in depressed and healthy older adults (Goldstein *et al.*, 1992). The RBMT and TEA were the best predictors of general functional impairment in multiple sclerosis (MS) patients compared to traditional cognitive tests (Higginson *et al.*, 2000). The TEA was also efficient in detecting a decline in the attentional abilities of healthy older adults (Robertson *et al.*, 1994). The BADS was substantially correlated with the everyday executive functioning (Evans, Chua, McKenna, & Wilson, 1997) and general cognitive functioning (Norris & Tate, 2000) of neurological patients and healthy individuals. Lastly, the CAMPROMPT was the only significant predictor of the vocational abilities (e.g., returning to work, withdrawal, or impaired performance) of MS patients amongst other tests such as the Auditory Consonant Trigrams test, the Zoo Map test, and the Screening Examination for Cognitive Impairment (Honan *et al.*, 2015).

Therefore, the above-mentioned neuropsychological tests have achieved a certain level of ecological validity in terms of both veridicality and verisimilitude. However, these neuropsychological tests suffer from several limitations. As discussed above, verisimilitude is the degree that the neuropsychological test replicates the environment (e.g., a supermarket, a living-room, a kitchen, a shopping mall, or a car) and procedures that a specific behaviour (e.g., shopping, cooking, taking medications, driving, or cleaning) will take place (Goldstein, 1996; Rabin, Burton, & Barr, 2007; Parsons, 2015). Since the preliminary discussions on the verisimilitude of neuropsychological testing, the emphasis has been put on the inability of the available technologies (e.g., paper-and-pencil, two-dimensional interfaces, as well as simple and static stimuli) to provide an adequate degree of verisimilitude (Goldstein, 1996). Unfortunately, two decades after, most of the neuropsychological assessments incorporate the same outdated technologies (e.g., paper-and-pencil, two-dimensional interfaces, static stimuli) which are not capable of providing a high degree of verisimilitude and inform on real-world functioning (Parsons, 2015; Parsons *et al.*, 2018; Rabin *et al.*, 2007). Hence, regardless that these paper-and-pencil tests present a higher degree of verisimilitude compared to traditional tests, their form (i.e., paper-and-pencil) allows them to incorporate only simple and static stimuli within a highly controlled environment, which substantially diverges from the cognitive demands and complexity of real-life situations (Parsons, 2015; Parsons *et al.*, 2018; Rand *et al.*, 2009).

1.4.4. Real world tasks

Another approach to studying everyday cognitive functioning is the design of neuropsychological tasks in real-world settings. For example, the participant is required to perform a series of errands in a shopping centre or a pedestrianized street (e.g., Garden, Phillips, & MacPherson, 2001; Shallice & Burgess, 1991). One of the most extensively utilised real-world tests is the Multiple Errands Test (MET) developed by Shallice and Burgess (1991). The MET requires the examinee to perform a series of relatively simple open-ended tasks (e.g., shopping for specific products, noting down specific information, or commuting to a specific location) and respect a set of simple rules (e.g., when you enter a shop, you must buy at least one item). The examiner observes the examinee performing the tasks and notes down the quantity and the type of errors (e.g., breaking a rule, or omitting a task). Regarding real-world tasks like MET, the relevant literature has shown that performance on real-world tasks (e.g., household chores) is significantly associated with self-ratings of IADL and independence questionnaires (Weakley, Weakley, & Schmitter-Edgecombe, 2019), and the findings of these studies may reliably generalise the everyday functionality of the assessed population (e.g., patients with specific traumatic brain injuries; Bottari, Dassa, Rainville, & Dutil, 2010; Bottari, Shun, Le Dorze, Gosselin, & Dawson, 2014).

These real-world tasks facilitate a direct examination of everyday cognitive functioning which benefits from both verisimilitude and veridicality. Nonetheless, the real-word tasks cannot be standardized, which prevents their implementation in other clinics or laboratories (Elkind, Rubin, Rosenthal, Skoff, & Prather, 2001; Logie *et al.*, 2011; Parsons, 2015; Rand *et al.*, 2009). Moreover, they may not be feasible for some individuals in challenging populations (e.g., psychiatric patients, stroke patients with paresis or paralysis), and they require participant transport and consent from local businesses (Elkind *et al.*, 2001; Logie *et al.*, 2011; Parsons, 2015; Rand *et al.*, 2009). Finally, real-world tasks are time-consuming and expensive, and importantly, they lack experimental control over the external situation (Elkind *et al.*, 2001; Logie *et al.*, 2011; Parsons, 2015; Rand *et al.*, 2009).

1.4.5. Non-immersive virtual reality methods

More recent attempts to achieve ecological validity in terms of verisimilitude involve the utilization of technological mediums. For example, some studies have used video recordings of real-world locations in an attempt to simulate everyday settings and situations (Farrimond, Knight, & Titov, 2006; McGeorge *et al.*, 2001; Paraskevaides *et al.*, 2010). However, non-immersive virtual reality has been implemented more frequently to simulate real-life tasks (Farrimond, Knight, & Titov, 2006; McGeorge *et al.*, 2001; Paraskevaides *et al.*, 2010). There are several non-immersive virtual reality tests such as Edinburgh Virtual Errands Test (EVET; Logie, Trawley, & Law, 2011), the Jansari Assessment of Executive Function (Jansari *et al.*, 2014), the Virtual Multiple Errands Test (VMET) within the Virtual Mall (VMall; Rand *et al.*, 2009) and the Virtual Reality Shopping Task (Canty *et al.*, 2014). These tests have been found to be cost-effective, require less administration time than paper-and-pencil tests, provide an adequate level of experimental control, and can be easily be adapted for other clinical or research settings (Parsons, McMahan, & Kane, 2018; Werner & Korczyn, 2012; Zygouris & Tsolaki, 2015). Furthermore, non-immersive virtual reality tests may benefit from automated scoring and standardized administration, which permit clinicians and researchers to administer these tests with only limited training (Parsons *et al.*, 2018; Werner & Korczyn, 2012; Zygouris & Tsolaki, 2015). Finally, some non-immersive virtual reality tests also offer shorter versions for the assessment of particular cognitive functions (e.g., selective visual attention; Parsons *et al.*, 2018; Werner & Korczyn, 2012; Zygouris & Tsolaki, 2015).

Non-immersive virtual reality tests appear more efficient in achieving ecological validity than paper-and-pencil and two-dimensional computerized tasks (Parsons, 2015; Parsons *et al.*, 2018). However, the utilization of non-immersive virtual reality technology may result in adverse symptoms and effects such as nausea, dizziness, disorientation, fatigue, or instability, which are frequently referred as virtual reality induced symptoms and effects (VRISE; Bohil, Alicea, & Biocca, 2011; de Franca & Soares, 2017; Palmisano, Mursic, & Kim, 2017). Also, non-immersive virtual tests have user interfaces (e.g., three-dimensional environments displayed on a screen) and procedures (e.g., using a keyboard and a mouse to interact with the virtual environment) which have been found to be challenging for individuals without a gaming background (Parsons *et al.*, 2018; Zaidi, Duthie, Carr, & Maksoud, 2018), especially for older adults and clinical populations such as individuals with mild cognitive impairment or Alzheimer's disease (Werner & Korczyn, 2012; Zygouris & Tsolaki, 2015). Notably, the gaming ability of the individual substantially modulates the performance on non-immersive virtual reality applications (Zaidi *et al.*, 2018).

1.5. Immersive virtual reality methods

1.5.1. Advantages of immersive virtual reality methods

In recent years, immersive virtual reality technology has attracted attention, demonstrating its utility and benefits in the field of cognitive neuroscience and neuropsychology (Rizzo, Schultheis, Kerns, & Mateer, 2004; Bohil *et al.*, 2011; Parsons, 2015). In contrast to the approaches discussed above, immersive virtual reality offers the utilisation of dynamic stimuli, naturalistic interactions within an ecologically valid 360° environment, and a high degree of experimental control which enables the collection of advanced cognitive and behavioural data (Rizzo *et al.*, 2004; Bohil *et al.*, 2011; Parsons, 2015). Furthermore, immersive virtual reality systems can be implemented in conjunction with non-invasive neuroimaging techniques (Bohil *et al.*, 2011; Parsons, 2015). Immersive virtual reality research and clinical software have been effective in the assessment of cognitive and affective functions, as well as clinical conditions (e.g., social stress disorders) which require ecological validity (Rizzo *et al.*, 2004; Parsons, 2015) for their assessment, rehabilitation and treatment (e.g., post-traumatic stress disorder) (Rizzo *et al.*, 2004; Bohil *et al.*, 2011).

While immersive virtual reality tests maintain the same advantages as non-immersive tests, they may also overcome the shortcomings of non-immersive virtual reality (Rizzo *et al.*, 2004; Bohil *et al.*, 2011; Parsons, 2015; Teo *et al.*, 2016). For example, immersive virtual reality has been found to provide deeper immersion in the virtual environments than non-immersive virtual reality systems (Weech, Kenny, & Barnett-Cowan, 2019). Deeper immersion has been found to induce substantially fewer and less intense adverse VRISE (Weech *et al.*, 2019). Importantly, in immersive virtual reality environments, individuals without gaming experience have been found to perform better, and comparable to individuals with gaming experience (Zaidi *et al.*, 2018). In immersive virtual reality systems, gaming ability does not have an effect on performance due to the first-person perspective and ergonomic/naturalistic interactions that are proximal to real-life settings and actions (Zaidi *et al.*, 2018).

1.5.2. Limitations of immersive virtual reality methods

Despite the important advantages of immersive virtual reality systems, researchers and clinicians have reported caveats with their implementation, particularly when head mounted display (HMD) systems are utilized (Davis, Nesbitt, & Nalivaiko, 2015; de Franca & Soares, 2017; Palmisano *et al.*, 2017; Sharples, Cobb, Moody, & Wilson, 2008). A predominant concern is the presence of adverse VRISE (Davis *et al.*, 2015; de Franca & Soares, 2017; Palmisano *et al.*, 2017; Sharples *et al.*, 2008), which may risk the health and safety of participants or patients (Kane & Parsons, 2017; Parsons *et al.*, 2018), and consequently raises ethical considerations for the adoption of immersive virtual reality HMDs as research and clinical tools.

Additionally, the presence of intense VRISE appears to modulate a substantial decline in reaction times and the overall cognitive performance of individuals (Mittelstaedt, Wacker, & Stelling, 2019; Nalivaiko, Davis, Blackmore, Vakulin, & Nesbitt, 2015; Nesbitt, Davis, Blackmore, & Nalivaiko, 2017; Plant, 2016; Plant & Turner, 2009). Also, VRISE seem to increase body temperature and heart rates (Nalivaiko *et al.*, 2015). Regarding brain activity and connectivity, the presence of intense VRISE substantially increases cerebral blood flow and oxyhaemoglobin concentration (Gavgani et al, 2018), the power of brain signals (Arafat, Ferdous, & Quarles, 2018), and the connectivity between stimulus response brain regions (e.g., prefrontal cortex) and nausea-processing brain regions (e.g., temporal and occipital lobes; Toschi *et al.*, 2017). Therefore, VRISE may confound the data deriving from neuropsychological (e.g., cognitive tests and screens), physiological (e.g., electromyography or thermal cameras), and neuroimaging methods (e.g., electroencephalography or magnetic resonance imaging).

Lastly, VRISE are evaluated using questionnaires such as the Simulator Sickness Questionnaire (SSQ; Kennedy, Lane, Berbaum, & Lilienthal, 1993) and the Virtual Reality Sickness Questionnaire (VRSQ; Kim, Park, Choi, & Choe, 2018). The SSQ is the most frequently used questionnaire. The SSQ includes 16 questions pertinent to the symptoms of simulator sickness (e.g., general discomfort, fatigue, headache, salivation, nausea, dizziness, eyestrain, blurred vision, vertigo, and burping; Kennedy *et al.*, 1993). The responses are assigned numbers indicating either the presence or the severity of simulator sickness symptoms (i.e., 0 – None, 1 – Slight, 2 – Moderate, and 3 – Severe; Kennedy *et al.*, 1993). However, the SSQ is specific to the symptoms pertinent to simulator sickness (Kennedy *et al.*, 1993), while simulator sickness symptomatology does not appear similar to VRISE (Stanney, Kennedy, & Drexler, 1997). Specifically, Stanney and collaborators (1997) clustered the 16 items of the SSQ into three main categories: 1) Nausea; 2) Oculomotor; 3) Disorientation. However, several of the SSQ items fall under of more than one category. For example, vertigo falls under both nausea and disorientation. Stanney et al. (1997) found that the most frequent symptoms related to VRISE were those related to the categories of nausea and disorientation, while the main symptoms related to simulator sickness fall under the oculomotor category. Also, the intensity of VRISE symptoms were substantially greater than the simulator sickness symptoms (Stanney *et al.*, 1997).

The necessity for a tool which is specific to VRISE, led to the development of the VRSQ (Kim *et al.*, 2018). The development of the VRSQ was based on the SSQ, where the researchers administered the SSQ and, by using an exploratory factor analysis, attempted to isolate the items of the SSQ which are pertinent to VRISE (Kim *et al.*, 2018). However, only items pertinent to oculomotor and disorientation categories of the SSQ displayed an adequate load to their factors, and all the items pertinent to nausea (i.e., 7 out of the 16 items of SSQ) were rejected (Kim *et al.*, 2018). Consequently, the VRSQ includes only 9 of the 16 items of the SSQ which exclusively relate to oculomotor and disorientation categories, while the most frequent VRISE relate to the nausea category (Bohil *et al.*, 2011; de França and Soares, 2017; Palmisano *et al.*, 2017; Sharples *et al.*, 2008; Stanney *et al.*, 1997). Hence, since the VRSQ has important omissions and does not encompass the main symptoms of VRISE, there is still a need for a questionnaire which measures the intensity of main VRISE.

1.5.2. Potential solutions

The physiological reasons for VRISE are predominantly due to oculomotor and vestibular discrepancies (Davis *et al.*, 2015; de Franca & Soares, 2017; Palmisano *et al.*, 2017; Sharples *et al.*, 2008). For example, there may be a discrepancy between what is being perceived through the oculomotor (optic nerve) sensor and what is being sensed via the rest of the afferent nerves in the human body (Davis *et al.*, 2015; de Franca & Soares, 2017; Palmisano *et al.*, 2017; Sharples *et al.*, 2008). Nevertheless, oculomotor and vestibular discrepancies are predominantly induced by hardware and software inadequacies, such as the type of display screen, resolution and refresh rate of the image, the size of the field of view as well as non-ergonomic movements and interactions with the virtual environment (de Franca & Soares, 2017; Palmisano *et al.*, 2017).

The reports of VRISE derive from studies which have implemented obsolete immersive virtual reality HMDs, while the last decade immersive virtual reality HMDs have substantially evolved (de Franca & Soares, 2017; Palmisano *et al.*, 2017). For example, market dominant companies like HTC and Facebook released their immersive virtual reality products in 2016. The HTC Vive and Facebook's Oculus Rift are new generation immersive virtual reality HMDs which have better image resolution and refresh rates, a faster and more accurate motion tracking system, and more ergonomic controllers than the old generation immersive virtual reality HMDs (Borrego, Latorre, Alcañiz, & Llorens, 2018). These HMDs meet the health and safety standards of these companies, and they do not seem to induce VRISE (Borrego *et al.*, 2018). However, immersive virtual reality software features should also be considered for the avoidance or mitigation of VRISE (de Franca & Soares, 2017; Palmisano *et al.*, 2017). Also, as discussed above, the development of a questionnaire to assess VRISE intensity is required. However, as discussed above, the quality of the software features implicates in the presence and intensity of VRISE. Therefore, this questionnaire may be developed to assess both the quality of software features and the VRISE intensity.

1.6. Interim summary

In summary, everyday functioning is facilitated by cognitive functions such as prospective memory, episodic memory, attention, and executive functions. Prospective memory functioning, which appears crucial in everyday functioning, is also mediated by episodic memory, attentional processes, and executive functions. Thus, the assessment of prospective memory and these associated cognitive functions plays an important role in understanding the everyday cognitive functioning of an individual. In order to achieve this, the neuropsychological assessment of everyday cognitive functions should be ecologically valid, which is achieved by adopting a veridicality and/or verisimilitude approach, where verisimilitude appears to be more effective than veridicality. Traditional paper-and-pencil tests and computerized tasks lack an adequate level of verisimilitude. Those paper-and-pencil tests with verisimilitude appear to be effective and have veridicality but they still do not resemble the complexity and cognitive demands of real-life situations. Real-world tasks are not feasible for implementation in clinics and laboratories albeit their efficiency in assessing everyday functioning. Non-immersive virtual reality tests appear to substantially improve the verisimilitude of neuropsychological assessment. Nonetheless, they also suffer from considerable limitations such as the confounding effects of VRISE and the gaming ability of individuals. In immersive virtual reality applications, which seem capable of closely resembling everyday settings and situations, gaming ability does not seem to play a significant role, and adverse VRISE are thought to be mitigated by recent technological advancements. However, the VRISE may still have a confounding effect on neuroscientific and/or neuropsychological data, while a tool for measuring both VRISE and immersive virtual reality software features is required to assess whether this is the case or not.

1.7. Objectives and overall scope

This thesis aims to address the problem of ecological validity in neuropsychological assessment, especially regarding the assessment of cognitive functions which are central to everyday functioning. This thesis also endeavours to overcome the shortcomings associated with the implementation of immersive virtual reality. Specifically, I aim to develop and validate the Virtual Reality Everyday Assessment Lab (VR-EAL), the first immersive virtual reality neuropsychological battery with enhanced ecological validity of everyday cognitive functions such as prospective memory, episodic memory, visual attention, visuospatial attention, auditory (bi-aural) attention, planning, and multitasking. It is also hypothesised that the VR-EAL will not induce intense VRISE. For this reason, the technological aetiologies of VRISE, both in terms of hardware and software characteristics, are carefully examined. The identification of appropriate hardware (i.e., HMDs) for the avoidance of VRISE indicates which hardware the VR-EAL should be compatible with. Also, the identification of software features which mitigate the intensity of VRISE will inform on which software characteristics the VR-EAL should incorporate. Finally, since the VR-EAL is developed for both clinical and research purposes, VR-EAL's validity and utility as neuropsychological tool is assessed, as well as VR-EAL's potency to contribute to the understanding of essential everyday cognitive functions such as prospective memory.

Firstly, Chapter 2 (published, Frontiers in Human Neuroscience) provides a systematic literature review of the technological aetiologies of VRISE, and a meta-analysis in terms of VRISE frequency, participants' dropouts, generation of the HMD, and the type of interactions within the virtual environment of the neuroscientific and neuropsychological studies which had implemented immersive virtual reality HMDs and software. Chapter 3 (published, Frontiers in Human Neuroscience) describes the validation of the Virtual Reality Neuroscience Questionnaire (VRNQ), which measures both VRISE and software quality, and explores the maximum duration of the virtual reality sessions without the presence of VRISE. Chapter 4 (published, Frontiers in Computer Science) presents and discusses the development of the VR-EAL and investigates whether VR-EAL induces VRISE. Chapter 5 (in press, Journal of the International Neuropsychological Society) discusses a comparison in terms of testing experience (i.e., pleasantness) and verisimilitude, as well as the convergent, construct, and ecological validity of VR-EAL against established ecological valid tests such as CAMPROMPT, RBMT-III, TEA, and BADS. Chapter 6 (submitted, Memory and Cognition) discusses the differences in the performance on the VR-EAL's focal event-based, non-focal event-based, and time-based prospective memory tasks, as well as the predictors (i.e., cognitive functions) of focal event-based, non-focal event-based, and time-based prospective memory scores of VR-EAL. Finally, Chapter 7 provides an all-inclusive discussion of the findings of this thesis.

Chapter 2: Technological Competence is a Precondition for Effective Implementation of Virtual Reality Head Mounted Displays in Human Neuroscience: A Technological Review and Meta-analysis

---

In this chapter, the hardware and software features of immersive virtual reality (VR) head-mounted display (HMD) systems, which contribute to the mitigation or avoidance of virtual reality induced symptoms and effects (VRISE), are explored by conducting a technological systematic review. Also, a meta-analysis of cognitive neuroscience and neuropsychology studies which had implemented VR HMDs is conducted to investigate the frequency of VRISE and the size of dropouts, when using appropriate or inappropriate VR HMDs and software. The findings of this chapter may assist with the identification of the appropriate VR HMDs which VR-EAL should be compatible with, as well as the software features that the VR-EAL should incorporate to ensure a mitigation or avoidance of VRISE.

This chapter has been published in the Frontiers in Human Neuroscience Frontiers: Kourtesis, P., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2019). Technological competence is a precondition for effective implementation of virtual reality head mounted displays in human neuroscience: a technological review and meta-analysis. *Frontiers in Human Neuroscience*, 13, 342. doi: https://doi.org/10.3389/fnhum.2019.00342

---

frontiers in Human Neuroscience

SYSTEMATIC REVIEW
published: 02 October 2019
doi: 10.3389/fnhum.2019.00342

Technological Competence Is a Pre-condition for Effective Implementation of Virtual Reality Head Mounted Displays in Human Neuroscience: A Technological Review and Meta-Analysis

Panagiotis Kourtesis [1,2,3,4]*, *Simona Collina* [3,4], *Leonidas A. A. Doumas* [2] and *Sarah E. MacPherson* [1,2]

[1] *Human Cognitive Neuroscience, Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom,* [2] *Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom,* [3] *Lab of Experimental Psychology, Suor Orsola Benincasa University of Naples, Naples, Italy,* [4] *Interdepartmental Centre for Planning and Research "Scienza Nuova", Suor Orsola Benincasa University of Naples, Naples, Italy*

OPEN ACCESS

Edited by:
Valerio Rizzo,
University of Palermo, Italy

Reviewed by:
Camila Rosa De Oliveira,
Faculdade Meridional (IMED), Brazil
Uri Maoz,
Chapman University, United States

***Correspondence:**
Panagiotis Kourtesis
pkourtes@exseed.ed.ac.uk

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *27 June 2019*
Accepted: *18 September 2019*
Published: *02 October 2019*

Citation:
Kourtesis P, Collina S, Doumas LAA and MacPherson SE (2019) Technological Competence Is a Pre-condition for Effective Implementation of Virtual Reality Head Mounted Displays in Human Neuroscience: A Technological Review and Meta-Analysis. Front. Hum. Neurosci. 13:342. doi: 10.3389/fnhum.2019.00342

Immersive virtual reality (VR) emerges as a promising research and clinical tool. However, several studies suggest that VR induced adverse symptoms and effects (VRISE) may undermine the health and safety standards, and the reliability of the scientific results. In the current literature review, the technical reasons for the adverse symptomatology are investigated to provide suggestions and technological knowledge for the implementation of VR head-mounted display (HMD) systems in cognitive neuroscience. The technological systematic literature indicated features pertinent to display, sound, motion tracking, navigation, ergonomic interactions, user experience, and computer hardware that should be considered by the researchers. Subsequently, a meta-analysis of 44 neuroscientific or neuropsychological studies involving VR HMD systems was performed. The meta-analysis of the VR studies demonstrated that new generation HMDs induced significantly less VRISE and marginally fewer dropouts. Importantly, the commercial versions of the new generation HMDs with ergonomic interactions had zero incidents of adverse symptomatology and dropouts. HMDs equivalent to or greater than the commercial versions of contemporary HMDs accompanied with ergonomic interactions are suitable for implementation in cognitive neuroscience. In conclusion, researchers' technological competency, along with meticulous methods and reports pertinent to software, hardware, and VRISE, are paramount to ensure the health and safety standards and the reliability of neuroscientific results.

Keywords: virtual reality, VRISE, HMD, cybersickness, neuroscience, neuropsychology, psychology, VR

INTRODUCTION

In recent years, virtual reality (VR) technology has attracted attention, demonstrating its utility and potency in the field of neuroscience and neuropsychology (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015). Traditional approaches in human neuroscience involve the utilization of static and simple stimuli which arguably lack ecological validity (Parsons, 2015). VR offers the usage of dynamic stimuli and interactions with a high degree of control within an ecologically valid environment which enables the collection of advanced cognitive and behavioral data (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015). VR can be combined with non-invasive imaging techniques (Bohil et al., 2011; Parsons, 2015) and has been effective in the assessment of cognitive and affective functions and clinical conditions (e.g., social stress disorders) which require ecological validity (Rizzo et al., 2004; Parsons, 2015) for their assessment, rehabilitation and treatment (e.g., post-traumatic stress disorder) (Rizzo et al., 2004; Bohil et al., 2011).

However, researchers and clinicians have reported caveats with the implementation of immersive VR interventions and assessments, particularly when head mounted display (HMD) systems are utilized (Sharples et al., 2008; Davis et al., 2015; de França and Soares, 2017; Palmisano et al., 2017). A predominant concern is the presence of adverse physiological symptoms (i.e., cyber/simulation-sickness which includes nausea, disorientation, instability, dizziness, and fatigue). These undesirable effects are categorized as VR Induced Symptoms and Effects (VRISE) (Sharples et al., 2008; Davis et al., 2015; de França and Soares, 2017; Palmisano et al., 2017), and are evaluated by using questionnaires such as the Simulator Sickness Questionnaire (Kennedy et al., 1993) and the Virtual Reality Sickness Questionnaire (Kim et al., 2018).

VRISE may risk the health and safety of participants or patients (Kane and Parsons, 2017; Parsons et al., 2018), which raises ethical considerations for the adoption of VR HMDs as research and clinical tools. Additionally, the presence of VRISE has modulated substantial decline in reaction times and overall cognitive performance (Plant and Turner, 2009; Nalivaiko et al., 2015; Plant, 2016; Nesbitt et al., 2017; Mittelstaedt et al., 2018), as well as increasing body temperature and heart rates (Nalivaiko et al., 2015). Also, the presence of VRISE robustly increases cerebral blood flow and oxyhemoglobin concentration (Gavgani et al., 2018), the power of brain signals (Arafat et al., 2018), and the connectivity between stimulus response brain regions and nausea-processing brain regions (Toschi et al., 2017). Thus, VRISE could be considered confounding variables, which significantly undermine the reliability of neuropsychological, physiological, and neuroimaging data.

VRISE are predominantly mediated by an oculomotor discrepancy between what is being perceived through the oculomotor (optic nerve) sensor and what is being sensed via the rest of the afferent nerves in the human body (Sharples et al., 2008; Davis et al., 2015; de França and Soares, 2017; Palmisano et al., 2017). Nevertheless, technologically speaking, VRISE are derivatives of hardware and software inadequacies, i.e., the type of display screen, resolution, and refresh rate of the image, the size of the field of view (FOV) as well as non-ergonomic movements within an interaction in the virtual environment (VE; de França and Soares, 2017; Palmisano et al., 2017). Notably, VR HMDs have substantially evolved during the last two decades. Important differences may be seen between the HMDs released before 2013 (old generation) and those released from 2013 onwards (new generation). While the last old generation HMD was released in 2001 (i.e., nVisor SX111), the year 2013 is used to distinguish between old and new generation HMDs, since it is the year that the first new generation HMD prototype (i.e., Oculus Development Kit 1) was released. This systematic review attempts to clarify the technological etiologies of VRISE and provide pertinent suggestions for the implementation of VR HMDs in cognitive neuroscience and neuropsychology. In addition, a meta-analysis of the neuroscience studies that have implemented VR HMDs will be conducted to elucidate the frequency of VRISE and dropout rates as per the VR HMD generation.

TECHNOLOGICAL SYSTEMATIC REVIEW

In **Table 1**, a glossary of the key terms and concepts is provided to assist with comprehension of the utilized terminology. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines using a decremental stepwise method to perform the literature review (see **Figure 1**). The selected papers and book chapters included an explicit explanation and discussion of VRISE and users' experiences pertinent with the specified technological features of the VR hardware and software. Digital databases specialized in technologies were used: (1) IEEE Xplore Digital Library; (2) ACM Digital Library; (3) ScienceDirect; (4) MIT CogNet; and (5) Scopus. Two categories of keywords were used, where each category had three or more keywords and each paper had to include at least one keyword from each category in the main body of the text. The categories were: (1) "virtual reality" OR "immersive virtual reality" OR "head-mounted display;" AND (2) "VRISE" OR "motion sickness" OR "cyber sickness" OR "simulation sickness." Finally, the extracted information from the identified papers was clustered together under common features (i.e., display, sound, motion tracking, navigation, ergonomic interactions, user experience, and computer hardware).

Technological Etiologies of VRISE

Display

VR HMDs use the following three types of screens: Cathode Ray Tubes (CRT); Liquid Crystal Display (LCD); and organic light emitting diode (OLED). LCD screens replaced CRT ones due to VRISE (Costello, 1997). LCD, in comparison to CRT, alleviated the probability of visual complications and physical burdens (e.g., fatigue) (Costello, 1997). However, the suitability of LCD was challenged by the emergence of OLED screens. While old generation VR HMDs mainly utilize LCD screens (Costello, 1997), the commercial versions of new generation VR HMDs predominantly use OLED screens (Kim J. W. et al., 2017). The OLED screens have been found to be

TABLE 1 | Glossary of key terms and concepts.

	Terms and concepts	**Explanation/definition**
Headsets	Head mounted display (HMD)	A display device which is worn on the head and provides an immersive virtual reality for the wearer
	Development kit (DK) HMD	A prototype device, which is utilized by the VR Software developers to develop VR software before the commercial version of an HMD. The DKs are not provided for general use
	Commercial version (CV) HMD	The final version of an HMD, which is dispersed to the market for general use
Display	Liquid crystal display (LCD)	A type of display/screen that uses the light-modulating properties of liquid crystals. Liquid crystals emit light indirectly, instead of using a reflector to produce images
	Organic light emitting diode (OLED)	A type of display/screen that uses an organic compound film that emits light in response to an electric current. OLEDs are used as displays in devices such as television screens, computer monitors, and smartphones
	Field of view (FOV)	The area captured by the display device. The size of the FOV and the size of the display device directly affect the quality of the image
	Refresh rate and frame rate	The refresh rate is the number of times that the hardware updates its display per second. It involves the repeated display of identical frames. The frame rate indicates the frequency that software can add new data to a display
	Resolution	The number of distinct pixels in each dimension displayed in a frame
Interactions	Motion tracking	The process of tracking the movement of objects or people. It is facilitated by motion sensors which detect the position of motion trackers embedded in devices (e.g., HMDs and 6DoF controllers)
	Controllers/Wands with 6 degrees of freedom (DoF)	Controllers which have 6DoF of movement in 3-D space on three directional axes (i.e., Forward-Back, Left-Right, Up-Down) and three rotational axes (i.e., Roll, Pitch, Yaw)
	Direct hand interaction	A motion tracking device (i.e., a motion sensor) which directly tracks hand movements
	Teleportation	A navigation system, which allows the user to be transferred to a new location in the virtual environment without physically moving in the real environment
	Ergonomic interactions	These resemble real-life interactions, which optimize user experience and overall VR system performance (see also Definition of Ergonomic Interactions). Ergonomic interactions are facilitated by and restricted to the capabilities of the VR hardware and software
	Virtual environment (VE)	A three-dimensional artificial environment which is displayed on a display device and allows the users to interact with it

better than LCD screens for general implementation in VR, because of their faster response times, lighter weight, and better color quality (Kim J. W. et al., 2017). OLED screens decrease the likelihood of VRISE and offer an improved VR display (Kim J. W. et al., 2017).

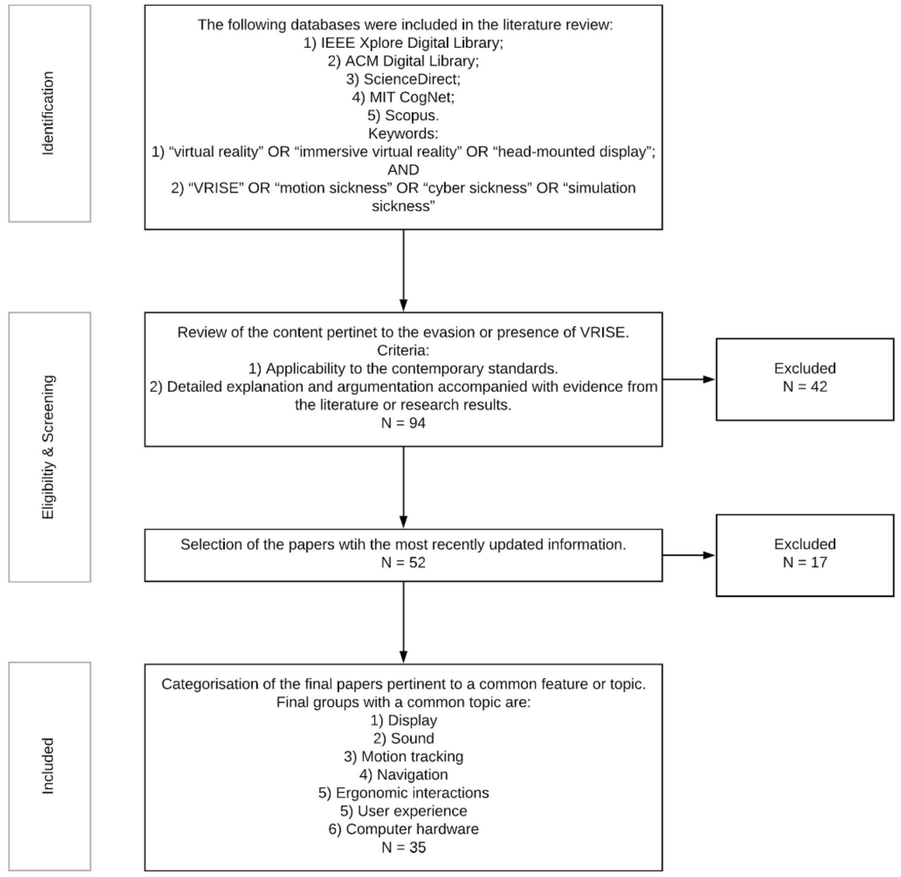
Three more factors related to display type are crucial for the avoidance of VRISE: the width of the FOV (Rakkolainen et al., 2016; Kim J. W. et al., 2017); the resolution of the image per eye (Hecht, 2016; Rakkolainen et al., 2016; Kim J. W. et al., 2017; Brennesholtz, 2018); and the latency of the images (frames per second) (Hecht, 2016; Rakkolainen et al., 2016; Kim J. W. et al., 2017; Brennesholtz, 2018). A wider FOV significantly decreases the chance of VRISE and increases the level of immersion (Rakkolainen et al., 2016; Kim J. W. et al., 2017). The canonical guidelines suggest a lowest threshold of 110° FOV (diagonal) (Hecht, 2016; Rakkolainen et al., 2016; Kim J. W. et al., 2017; Brennesholtz, 2018). In addition, an increased refresh rate and resolution alleviates the danger of discomfort or VRISE (Hecht, 2016; Rakkolainen et al., 2016; Kim J. W. et al., 2017; Brennesholtz, 2018). The refresh rate should be ≥75 Hz (i.e., ≥75 frames per s) (Goradia et al., 2014; Hecht, 2016; Brennesholtz, 2018), while the resolution is required to be higher than 960 × 1,080 sub-pixels per eye (Goradia et al., 2014).

Sound

A second important consideration for a user's experience in VR is the sound quality. The integration of spatialized sounds (e.g., ambient and feedback sounds) in the VE may increase the level of immersion, pleasantness of the experience, and successful navigation (Vorländer and Shinn-Cunningham, 2014), while they significantly decrease the likelihood of VRISE (Viirre et al., 2014). However, the volume and localization of sounds need to be optimized in terms of audio spatialization to ensure a user's experience is pleasant without adverse VRISE (Viirre et al., 2014; Vorländer and Shinn-Cunningham, 2014).

Motion Tracking

Motion tracking in VR is a pre-condition for naturalistic movement within an immersive VE (Slater and Wilbur, 1997; Stanney and Hale, 2014). Motion tracking allows the precise tracking of the user's physical body within the VE (i.e., it allows the computer to provide accurate environmental feedback, which modulates and consolidates the awareness of the position and movement of the user's body). This phenomenon is called proprioception or kinesthesia (Slater and Wilbur, 1997) and is linked with vestibular and oculomotor mediated VRISE (Slater and Wilbur, 1997; Plouzeau et al., 2015; Caputo et al., 2017). Hence, motion tracking should be adequately



The following databases were included in the literature review:
1) IEEE Xplore Digital Library;
2) ACM Digital Library;
3) ScienceDirect;
4) MIT CogNet;
5) Scopus.
Keywords:
1) “virtual reality” OR “immersive virtual reality” OR “head-mounted display”;
AND
2) “VRISE” OR “motion sickness” OR “cyber sickness” OR “simulation sickness”

Review of the content pertinet to the evasion or presence of VRISE.
Criteria:
1) Applicability to the contemporary standards.
2) Detailed explanation and argumentation accompanied with evidence from the literature or research results.
N = 94

Selection of the papers wtih the most recently updated information.
N = 52

Categorisation of the final papers pertinent to a common feature or topic.
Final groups with a common topic are:
1) Display
2) Sound
3) Motion tracking
4) Navigation
5) Ergonomic interactions
5) User experience
6) Computer hardware
N = 35

FIGURE 1 | Decremental stepwise method for the technological literature review.

rapid and accurate to facilitate ergonomic interactions in the VE (Caputo et al., 2017).

Navigation

A highly important factor in the quality of VR software and to avoid VRISE is the movement of the user in the VE (Porcino et al., 2017). New generation HMDs deliver an adequate play area for interactions to facilitate ecologically valid scenarios (Porcino et al., 2017; Borrego et al., 2018). However, there are restrictions in the size of the play area, which does not permit navigation solely by physical walking (Porcino et al., 2017; Borrego et al., 2018). Teleportation allows movement beyond the play area size and elicits a high-level of immersion and pleasant user experience, whilst alleviating VRISE (Bozgeyikli et al., 2016; Frommel et al., 2017; Porcino et al., 2017). In contrast, movement dependent on a touchpad, keyboard, or joystick results to high occurrences of VRISE (Bozgeyikli et al., 2016; Frommel et al., 2017; Porcino et al., 2017). Therefore, teleportation in conjunction with physical movement (i.e., free movement of the upper limbs and walking in a small-restricted area) is the most suitable method for movement in VR (Bozgeyikli et al., 2016;

TABLE 2 | Minimum hardware criteria: old and new generation VR HMDs.

Product	Generation	Resolution (per eye)	Display screen	Refresh rate	FOV (Diagonal)	Motion trackers and sensors (Type and quantity)
VFX 3D	Old	480 × 240	LCD	45 Hz	45°	–
VUZIX Wrap 1200	Old	852 × 480	LCD	60 Hz	35°	Unknown type (1), 3 magnetometers, 3 accelerometers, and 3 gyroscopes
eMagin Z800 3DVisor	Old	800 × 600	OLED	60 Hz	40°	–
nVisor SX111	Old	1,280 × 1,024	LCD	60 Hz	110°	–
Oculus rift development kit 1	New	640 × 800	LCD	60 Hz	110°	–
Oculus rift development kit 2	New	960 × 1,080	OLED	75 Hz	110°	–
Minimum hardware criteria for the avoidance of VRISE	NA	>960 × 1,080	OLED or LCD	≥75Hz	≥110°	Tracking should be adequately rapid and accurate to facilitate ergonomic interactions
Oculus rift commercial version	New	1,080 × 1,200	OLED	90 Hz	110°	Accelerometer, gyroscope, magnetometer, 360° constellation tracking camera
HTC VIVE commercial version	New	1,080 × 1,200	OLED	90 Hz	110°	Sensors (>70) including MEMS, magnetometer, gyroscope, accelerometer, and laser position sensors, lighthouse laser tracking system (2 base stations emitting pulsed InfraRed lasers), front-facing camera

MEMS, Microelectromechanical systems.

TABLE 3 | Criteria for suitable VR software in cognitive neuroscience and neuropsychology.

Domains	User experience	Game mechanics	In-game assistance	VRISE
Criteria	An adequate level of immersion	A suitable navigation system (e.g., Teleportation)	Digestible tutorials	Absence or insignificant presence of nausea
	Pleasant VR experience	Availability of physical movement	Helpful tutorials	Absence or insignificant presence of disorientation
	High quality graphics	Naturalistic picking/placing of items	Adequate duration of tutorials	Absence or insignificant presence of dizziness
	High quality sounds	Naturalistic use of items	Helpful in-game instructions	Absence or insignificant presence of fatigue
	Suitable hardware (HMD and computer)	Naturalistic 2-handed interaction	Helpful in-game prompts	Absence or insignificant presence of instability

Frommel et al., 2017; Porcino et al., 2017). Yet, there are additional factors such as external hardware (i.e., controllers and wands), which are needed to facilitate optimal ergonomic interactions in VR.

Ergonomic Interactions

Ergonomic and naturalistic interactions are essential to minimize the risk of VRISE, while non-ergonomic and non-naturalistic interactions increase the occurrence of them (Slater and Wilbur, 1997; Stanney and Hale, 2014; Plouzeau et al., 2015; Caputo et al., 2017; Porcino et al., 2017). Importantly, controllers, joysticks, and keyboards do not support ergonomic and naturalistic interactions in VR (Plouzeau et al., 2015; Bozgeyikli et al., 2016; Caputo et al., 2017; Frommel et al., 2017; Porcino et al., 2017; Sportillo et al., 2017; Figueiredo et al., 2018). Instead, wands with 6 degrees of freedom (DoF) of movement (e.g., Oculus Rift and HTC Vive wands), and realistic interfaces with direct hand interactions (e.g., Microsoft's Kinect) facilitate naturalistic and ergonomic interactions (Sportillo et al., 2017; Figueiredo et al., 2018). Both hardware systems facilitate easy familiarization with their controls and their utilization (Sportillo et al., 2017; Figueiredo et al., 2018). However, direct hand interactions are easier than 6DoF controllers-wands in terms of familiarization with their controls and efficiency (Sportillo et al., 2017; Figueiredo et al., 2018). Direct hand interactions were also found to offer more pleasant user experiences (Sportillo et al., 2017; Figueiredo et al., 2018), although, they are substantially less accurate than 6DoF controller-wands (Sportillo et al., 2017; Figueiredo et al., 2018).

User Experience

Notably, ergonomic interactions might be available to the user; however, the user is required to learn the necessary interactions and how the VE functions to facilitate a pleasant user experience (Gromala et al., 2016; Jerald et al., 2017; Brade et al., 2018). The inclusion of comprehensible tutorials where the user may spend an adequate amount of time acquiring the necessary skills (i.e., navigation, use and grab of items, two-handed interactions) and knowledge of the VE (i.e., how it reacts to your controls) is crucial (Gromala et al., 2016; Jerald et al., 2017; Brade et al., 2018). Additionally, in-game instructions and prompts should be offered to the user through interactions in the VE (e.g., directional arrows, non-player characters, signs, labels, ambient sounds, audio, and videos) (Gromala et al., 2016; Jerald et al., 2017; Brade et al., 2018).

Computer Hardware

The computer hardware (i.e., the processor, graphics card, sound card) should at least meet the minimum requirements of the VR software and HMD (Anthes et al., 2016). The performance of VR HMDs is analogous to the computing power and the quality of the hardware (Stanney and Hale, 2014; Anthes et al., 2016; Borrego et al., 2018). The processor, graphics card, sound card, and operating system (e.g., Windows) need to be considered and reported because they modulate the performance of the software (Plant and Turner, 2009; Plant, 2016; Kane and Parsons, 2017; Parsons et al., 2018). Research software developers and researchers are required to be technologically competent in order to opt for the appropriate hardware and software to achieve their research and/or clinical aims (Plant, 2016; Kane and Parsons, 2017; Parsons et al., 2018).

Conclusions

Based on the outcomes of the above technological review, VR HMDs should have a good quality display-screen (i.e., OLED or upgraded LCD), an adequate FOV (i.e., diagonal FOV $\geq$ 110°), adequate resolution per eye (i.e., resolution $> 960 \times 1{,}080$ sub-pixels per eye), and an adequate image refresh rate (i.e., refresh rate $\geq$ 75 Hz) to safeguard the health and safety of the participants and the reliability of the neuroscientific results (see **Table 2**). Also, the VR HMD should have external hardware which offers an adequate VR area, fast and accurate motion tracking, spatialized audio, and ergonomic interactions. The computer's processor, graphics card, and sound card should meet the minimum requirements of the VR software and HMD too. New generation VR HMDs appear to have all the necessary hardware characteristics (i.e., graphics, level of immersion, and sound) to be used in ecological valid research and clinical paradigms (Borrego et al., 2018; see **Table 2** for a comparison between old and new generation HMDs). New generation VR HMDs have the required hardware to support and produce high-quality spatialized sounds in VEs (Borrego et al., 2018). Additionally, new generation VR HMDs have integrated rapid and precise motion tracking which facilitates naturalistic and ergonomic interactions within the VE (Borrego et al., 2018).

Both the Oculus development kit (DK) 1 and DK2 do not meet the minimum hardware features highlighted by the technological review, despite being new generation VR HMDs (see **Table 2**). The DK1 has substantially lower resolution per eye and image refresh rates, while the DK2 has marginally acceptable refresh rates, yet a slightly lower resolution per eye. These DKs are not available for general use but are used by professional developers to produce beta (early) versions of their games or apps (Goradia et al., 2014; Suznjevic et al., 2017). Moreover, they were removed from the market after the release of the Oculus Rift CV. VR HMDs should have hardware characteristics equal to or better than the commercial versions (CV) of the Oculus Rift and HTC Vive in order to ensure the health and safety of the participants, as well as the reliability of the neuroscientific results (i.e., physiological, neuropsychological, and neuroimaging data). The researchers and clinicians should have the technological competence to choose an HMD which is equal to or greater than the CVs of the Oculus Rift and HTC Vive (e.g., Valve Index, HTC Vive Pro, Oculus Quest, Pimax VR, and StarVR).

However, the VR software's features are equally important. The VR software should include an ergonomic interaction and navigation system, as well as tutorials, in-game instructions, and prompts. A suitable navigation system should combine teleportation and physical movement, while ergonomic interactions should include those that simulate real-life interactions by using a direct hands system or 6DoF controllers. Also, the tutorials, in-game instructions, and prompts should be informative and easy to follow, especially for experimental or clinical purposes where users should be equally able to interact with the VE (Plant and Turner, 2009; Plant, 2016; Kane and Parsons, 2017; Parsons et al., 2018). The criteria for effective VR software are displayed in **Table 3**. These criteria should be met before implementing VR software for research and/or clinical purposes. Otherwise, researchers or clinicians may compromise the reliability of their study's results (Plant and Turner, 2009; Plant, 2016; Kane and Parsons, 2017; Parsons et al., 2018), and/or jeopardize the health and safety of their participants/patients (Kane and Parsons, 2017; Parsons et al., 2018).

The above features enable researchers or clinicians to administer a sophisticated and pleasant VR experience, which substantially alleviates or eradicates adverse VRISE. Therefore, the technological competency of neuroscientists and neuropsychologists is a pre-condition for the efficient adoption and implementation of innovative technologies like VR HMDs in cognitive neuroscience or neuropsychology.

META-ANALYSIS OF VR STUDIES IN COGNITIVE NEUROSCIENCE

Literature Research and Inclusion Criteria

We followed the PRISMA guidelines to conduct the literature research using a decremental approach, where the selection commenced with a relatively vast accumulation of abstracts and concluded with a diminished list of full papers that comprise standardized and detailed VR research paradigms. The procedure is described in **Figure 2**. The following databases were used for the literature research: (1) PsycInfo; (2) PsycArticles; (3) PubMed; and (4) Medline. Two categories of keywords were used, with three keywords in each category. The minimum threshold for each study was the inclusion of at least one keyword from each category in the main body of text. The keywords

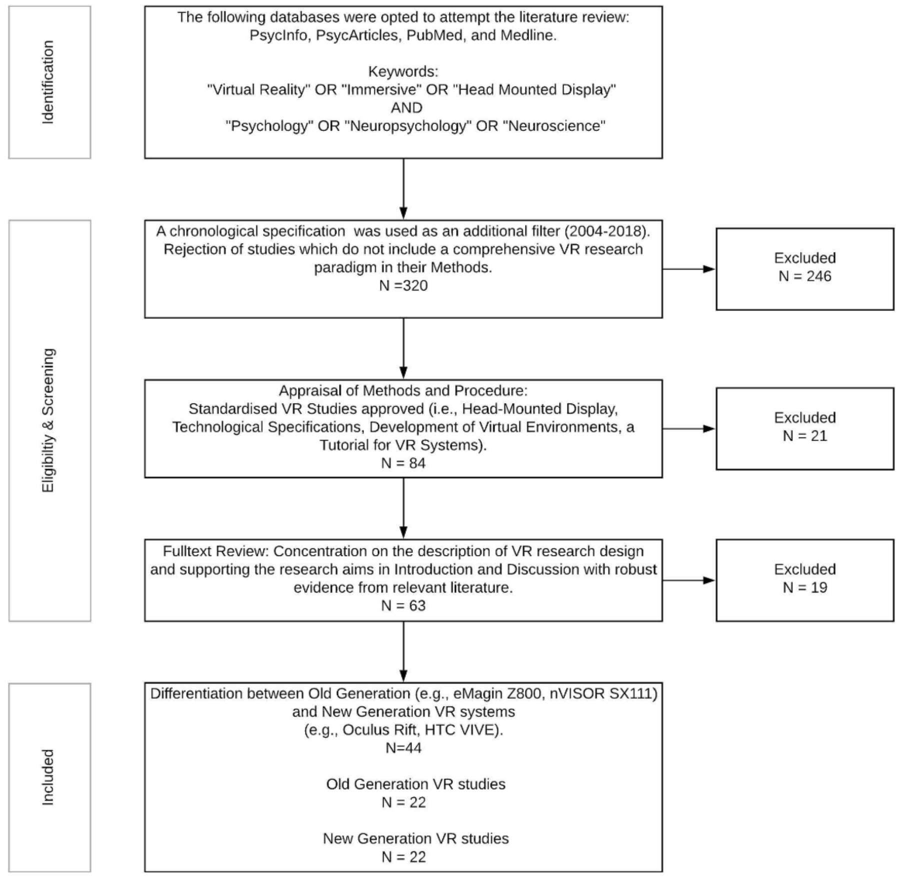


FIGURE 2 | Decremental stepwise method for the literature review of VR studies.

for each category were: (1) "virtual reality;" OR "Immersive;" OR "Head Mounted Display;" AND (2) "Psychology;" OR "Neuropsychology" OR "Neuroscience." Additional filters and criteria were: (1) chronological specification (2004 and later); and (2) a comprehensive description of the VR research methods in conjunction with the research aims and results. Finally, the selected studies were allocated into two groups according to the generation of the implemented VR HMD. Two tables display the studies that utilize old generation (**Table 4**) and new generation (**Table 5**) HMDs.

Data Collection and Coding

Target Variables

The principal aim of the meta-analysis was to measure the frequency of VRISE in neuroscience or psychology studies using a VR HMD. However, only six studies reported VRISE quantitatively (i.e., using a questionnaire). For this reason, we considered only the presence or absence of VRISE. The dichotomous VRISE variable (i.e., presence or absence of VRISE) was quantified (i.e., absent VRISE = 0; present VRISE = 1) to facilitate a comparison (i.e., Bayesian *t*-tests) between the studies

TABLE 4 | Neuroscience studies employing old generation VR HMDs.

References	Topic	HMD	Ergonomic interactions	Clinical condition	Age group	*N*	VRISE	Dropouts
			YES = 12 **NO = 10**				**YES = 14** **NO = 8**	**YES = 9** **NO = 13**
Kim et al. (2004)	Visuospatial functions	Eye-trek FMD-250W	YES	Brain injury	MA	52	YES	NO - 0
Moreau et al. (2006)	Executive functions	Eye-trek FMD-250W	YES	ADHD and autism	YA	22	YES	YES - 1
Botella et al. (2007)	Therapy (VRET)	V6 VR	NO	Panic disorder	YA and MA	46	YES	YES - 9
Matheis et al. (2007)	Memory	eMagin z800	YES	Brain injury	MA	40	NO	NO - 0
Parsons et al. (2007)	Executive functions	eMagin z800	NO	ADHD	C	20	YES	YES - 1
Banville et al. (2010)	Memory	eMagin z800	NO	Brain injury	YA	62	NO	NO - 0
Rizzo et al. (2010)	Therapy (VRET)	eMagin z800	NO	PTSD	YA	20	YES	YES - 5
Reger et al. (2011)	Therapy (VRET)	eMagin z800	NO	PTSD	YA	24	YES	YES - 6
Bioulac et al. (2012)	Executive functions	eMagin z800	YES	ADHD	YA	36	NO	NO - 0
Carlozzi et al. (2013)	Rehabilitation	eMagin z800	YES	Spinal cord injury	MA	54	YES	YES - 10
Meyerbroeker et al. (2013)	Therapy (VRET)	nVISOR SX111	YES	Agoraphobia	MA	55	YES	YES - 17
Parsons et al. (2013)	Attention assessment	eMagin Z800	YES	Healthy	YA	50	YES	NO - 0
Peck et al. (2013)	Racial biases	nVISOR SX111	YES	Healthy	YA	60	NO	NO - 0
Freeman et al. (2014)	Social cognition	nVISOR SX111	YES	Paranoia	YA	60	YES	NO - 0
Rothbaum et al. (2014)	Therapy (VRET)	eMagin Z800	NO	PTSD	YA and MA	156	NO	NO - 0
Veling et al. (2014)	Paranoid thoughts	eMagin Z800	NO	Psychosis	YA	41	YES	NO - 0
Hartanto et al. (2014)	Social stress	eMagin Z800	NO	Healthy	YA	54	NO	NO - 0
Gaggioli et al. (2014)	Stress levels	Vuzix Wrap 1200VR	NO	Healthy	MA	121	YES	NO - 0
Shiban et al. (2015)	Therapy (VRET)	eMagin Z800	NO	Arachnophobia	YA	58	YES	YES - 8
Freeman et al. (2016)	Therapy (VRET)	nVISOR SX111	YES	Persecutory delusions	MA	30	YES	YES - 1
Parsons and Carlew (2016)	Attention assessment	eMagin Z800	YES	Healthy	YA	50	NO	NO - 0
Parsons and Barnett (2017)	Attention assessment	eMagin Z800	YES	Healthy	YA and OA	89	NO	NO - 0

HMD, Head-Mounted Display; VRISE, VR induced adverse symptoms and effects; YA, Young Adults; MA, Middle-Aged Adults; OA, Older Adults; C, Children; VRET, VR Exposure Therapy; PTSD, Post-Traumatic Stress Disorder; ADHD, Attention Deficit Hyperactivity Disorder.

that used old generation HMDs, new generation DK HMDs, and new generation CV HMDs, as well as the examination of potential correlations with other variables (i.e., Bayesian Pearson's correlation analysis).

secondary aim of the meta-analysis was to inspect the dropout rates in neuroscience or psychology studies that used VR HMDs. However, as the vast majority of studies had no dropouts, studies with some dropouts (e.g., 3, 5, 6) were statistically considered as outliers. For this reason, we considered the existence of dropouts in each study. The dropout variable was dichotomized as presence = 1 and absence = 0. This dichotomous dropout variable was used to investigate whether using a certain generation HMD (i.e., old generation HMDs, new generation DK HMDs, or new generation CV HMDs) could increase/decrease the dropout size. We compared (i.e., Bayesian *t*-tests)

TABLE 5 | Neuroscience studies employing new generation VR HMDs.

References	Topic	HMD	Ergonomic interactions	Clinical condition	Age group	*N*	VRISE	Dropouts
			YES = 18 **NO = 4**				**YES = 4** **NO = 18**	**YES = 4** **NO = 18**
Foerster et al. (2016)	Attention assessment	Oculus DK2	NO	Healthy	YA	44	NO	YES - 2
Quinlivan et al. (2016)	Attention assessment	Oculus DK2	YES	Healthy	YA	40	NO	NO - 0
Kim A. et al. (2017)	VR presence	Oculus DK2	YES	PD	OA	33	NO	NO - 0
Montenegro and Argyriou (2017)	Memory, attention, executive functions	Oculus DK2	YES	AD (early stages)	OA	20	NO	NO - 0
Parsons and McMahan (2017)	Memory assessment	HTC vive	YES	Healthy	YA	103	NO	NO - 0
Kelly et al. (2017)	Spatial perception	HTC vive	YES	Healthy	YA	76	NO	NO - 0
Bourdin et al. (2017)	Fear of death	Oculus DK2	YES	Healthy	YA	36	NO	NO - 0
Hasler et al. (2017)	Racial bias	Oculus DK2	YES	Healthy	YA	36	NO	NO - 0
Mottelson and Hornbæk (2017)	Navigation, attention, B-P	HTC vive	YES	Healthy	YA and MA	31	NO	NO - 0
Rooney et al. (2017)	Social cognition	Oculus rift CV	YES	Healthy	YA and MA	103	NO	NO - 0
Zimmer et al. (2018)	Social stress	Oculus DK2	NO	Healthy	YA and MA	93	YES	YES - 5
Hsieh et al. (2018)	Spatial perception and navigation	HTC vive	YES	Healthy	YA	70	NO	NO - 0
Yeh et al. (2018)	Anxiety	HTC vive	YES	Healthy	YA	34	NO	NO - 0
Collins et al. (2018)	Psychoeducation on DBS	Oculus rift CV	YES	Movement disorder	OA	30	NO	NO - 0
Barberia et al. (2018)	Fear of death	Oculus DK2	YES	Healthy	YA	31	YES	YES - 1
Banakou et al. (2018)	Embodiment, cognition—IQ	HTC vive	YES	Healthy	YA	30	NO	NO - 0
Christou et al. (2018)	Motor-rehabilitation	HTC vive	YES	Stroke patients	YA and MA	29	NO	NO - 0
Gómez-Jordana et al. (2018)	Balance and walking rehabilitation	Oculus DK2	YES	PD	OA	22	NO	NO - 0
Lubetzky et al. (2018)	Sensory integration and balance	Oculus DK2	NO	Healthy	YA and MA	21	YES	NO - 0
Oagaz et al. (2018)	Memory assessment	HTC vive	YES	Healthy	YA	20	NO	NO - 0
George et al. (2018)	Working memory and attention assessment	HTC vive	YES	Healthy	YA	20	NO	NO - 0
Detez et al. (2019)	Gambling	HTC vive	NO	Healthy	YA and MA	60	YES	YES - 3

HMD, Head-Mounted Display; VRISE, VR induced adverse symptoms and effects; YA, Young Adults; MA, Middle-Aged Adults; OA, Older Adults; PD, Parkinson's disease; AD, Alzheimer's disease; DK, Development Kit; CV, Commercial Version; B-P, Body Perception; DBS, Deep Brain Stimulation.

the dropout rate across studies that used old generation HMDs, new generation DK HMDs, and new generation CV HMDs. We also inspected whether the dropout rates correlated with other variables by using Bayesian Pearson's correlation analysis.

Grouping Variables

We subdivided studies into groups based on the HMD generation they used. Hence, two groups of studies were created and compared by using Bayesian *t*-tests; the first group included studies that utilized old generation HMDs, while the second group included studies which utilized new generation HMDs (i.e., both DKs and CVs).

The new generation studies were further distinguished and compared by using Bayesian t-tests based on the type of new generation HMDs adopted (i.e., DK or CV). Two sub-groups were formed; the first group included studies that utilized DK HMDs, and the second group included studies that utilized CV HMDs.

Furthermore, the recency of the HMD technology was compared by using an ordinal variable where 1 indicated old generation HMDs, 2 indicated new generation DKs, and 3 indicated new generation CVs. This ordinal variable allowed us to inspect whether the HMD generation correlated with other variables by using Bayesian Pearson's correlation analysis.

Lastly, we considered the type of interactions, where the type of interactions were expressed in a binary form (i.e., non-ergonomic interactions = 0 and ergonomic interactions = 1). This allowed a comparison between the VR studies which had ergonomic interactions and the VR studies which had non-ergonomic interactions by using a Bayesian t-test. It also allowed us to inspect whether the interaction type correlated with other variables (i.e., Bayesian Pearson's correlation analysis).

Definition of Ergonomic Interactions

In line with the definition of ergonomic interactions in our technological review, we considered interactions to be ergonomic or non-ergonomic based on their proximity to real-life interactions. We provide some examples below to clarify our criteria:

> Example 1—Ergonomic Interaction: if the VR software required the participant to look around moving his or her head.
> Example 2—Non-Ergonomic Interaction: if the VR software required the participant to look around by using a joystick or mouse.
> Example 3—Ergonomic Interaction: if the VR software required the user to interact with objects (e.g., pushing a button, holding an item) in the VE or to navigate within the VE by using either 6DoF controllers or direct-hand interactions.
> Example 4—Non-Ergonomic Interaction: if the VR software required the user to interact with objects (e.g., pushing a button, holding an item) in the VE or to navigate within the VE by using a keyboard or joystick (e.g., Xbox controller).

Statistical Analyses

Bayesian statistics were preferred over null hypothesis significance testing (NHST). The Bayesian factor (BF_{10}) was therefore used instead of p-values for statistical inference, although we do report both BF_{10} and p-values. P-values measure the difference between the data and the null hypothesis (H_0) (e.g., the assumption of no difference or no effect), while the BF_{10} calibrates p-values by converting them into evidence in favor of the alternative hypothesis (H_1) over the H_0 (Cox and Donnelly, 2011; Bland, 2015; Held and Ott, 2018). BF_{10} is considered substantially more parsimonious than the p-value in evaluating the evidence against the H_0 (Cox and Donnelly, 2011; Bland, 2015; Held and Ott, 2018). Also, the difference between BF_{10} and the p-value in evaluating the evidence against H_0 is even greater in small sample sizes (Held and Ott, 2018). Bayesian Factor (BF_{10}) threshold ≥ 10 was set for statistical inference in all analyses, which indicates strong evidence in favor of the H_1 (Rouder and Morey, 2012; Wetzels and Wagenmakers, 2012; Marsman and Wagenmakers, 2017), and corresponds to a p-value < 0.01 (e.g., $BF_{10} = 10$) (Cox and Donnelly, 2011; Bland, 2015; Held and Ott, 2018). JASP software was used to perform the statistical analyses (JASP Team, 2018). Bayesian independent samples t-tests were conducted to investigate the difference in VRISE frequency and dropout occurrence between old and new generation HMDs, as well as between new generation DKs and CVs. A Bayesian Pearson's correlations analysis examined the possible statistical relationships amongst the HMD generations, VRISE presence, the type of interactions, and dropout occurrences.

Results

The Implementation of Old and New Generation HMDs in Cognitive Neuroscience

The studies that utilized old generation HMDs are displayed in **Table 4** and recruited 1,200 participants in total. Nine out of 22 studies examined stress disorders, 7 of these were VR exposure therapy (VRET) studies either for phobias or post-traumatic stress disorder (PTSD), while 2 studies attempted to assess stress levels in context (e.g., assessment of social stress during a job interview). In 9 studies, there were VR assessments of cognitive functions, 2 studies assessed memory, 3 studied attention, 3 examined executive functions, and 1 examined visuospatial ability. Two of the studies involved social cognition while only one involved paranoid thinking. Lastly, only one study provided rehabilitation sessions in VR for patients with spinal injuries. The targeted age groups were young adults in 18 studies, middle-aged adults in 8 studies, older adults in one study, and children in one study.

The studies that utilized new generation HMDs are displayed in **Table 5** and recruited 982 individuals in total. Specifically, 376 individuals were recruited in 10 studies where new generation DKs were used, while 606 individuals were recruited in 12 studies where new generation CVs were used. Nine out of the 22 studies attempted to assess cognitive functions (i.e., memory, attention, visuospatial ability, executive functions), 4 investigated anxiety disorders (i.e., fear of death, social stress, general anxiety disorder), 3 provided sensorimotor rehabilitation interventions, 3 studies examined the effects of presence in specific VEs, 2 assessed social cognition and 1 study offered a psychoeducational session to patients with motor-related disorders. Lastly, the targeted age groups were young adults in 18 studies, middle-aged adults in 6 studies, and older adults in 4 studies.

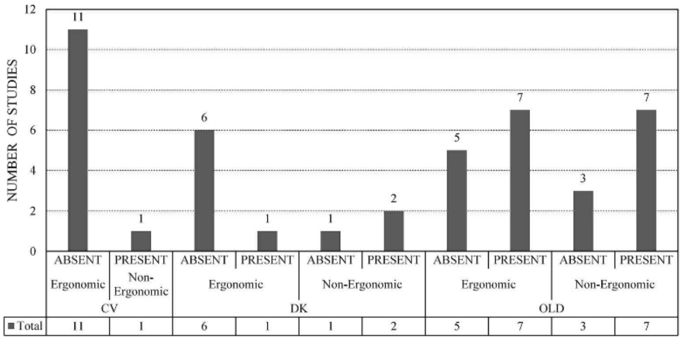


FIGURE 3 | VRISE per HMD generation and ergonomic interactions. ABSENT, Absence of VRISE; PRESENT, Presence of VRISE; OLD, Old Generation HMD; DK, New Generation HMD—Development Kit; CV, New Generation HMD—Commercial Version; Ergonomic, Ergonomic Interactions; Non-Ergonomic, Non-Ergonomic Interactions.

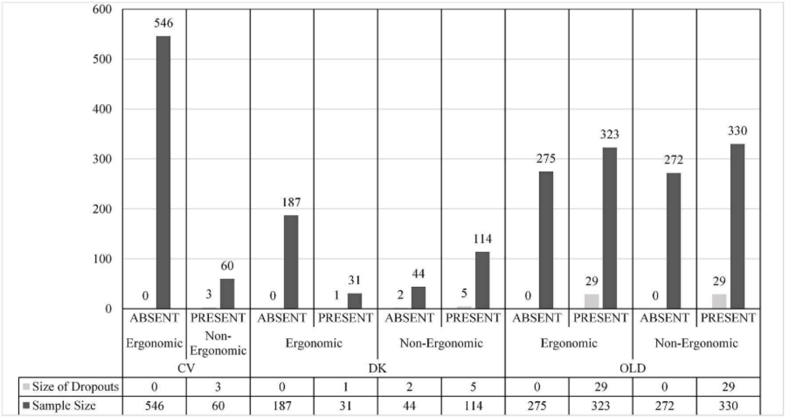


FIGURE 4 | Droupouts and sample size per HMD, VRISE presence, and ergonomic interactions. ABSENT, Absence of VRISE; PRESENT, Presence of VRISE; OLD, Old Generation HMD; DK, New Generation HMD—Development Kit; CV, New Generation HMD—Commercial Version; Ergonomic, Ergonomic Interactions; Non-Ergonomic, Non-Ergonomic Interactions.

Meta-Analysis

The descriptive statistics are presented in **Figures 3**, **4**. In **Figure 3**, the number of studies with VRISE are displayed according to their HMD generation and interaction type. In **Figure 4**, the dropouts and sample sizes are presented according to their HMD generation, VRISE presence, and interaction type. The presence of VRISE substantially becomes less frequent when new generation HMDs are implemented (**Figure 3**). In new generation HMDs, VRISE are present in only 4 out of 22 studies, while across 982 participants, there are only 11 dropouts. In contrast, in old generation HMDs, VRISE are present in 14 out of 22 studies, while in a total sample size of 1,200 participants, there are 58 dropouts.

In the 14 old generation HMDs studies where VRISE are present, half of them involved ergonomic interactions and the other half involved non-ergonomic interactions. Similarly, there is an equal distribution of dropouts (29 in each) between the old generation HMDs studies that had ergonomic and non-ergonomic interactions. When only old HMDs with ergonomic interactions are considered, VRISE are present in 7 out of 12 studies, while in an entire sample size of 598, the dropouts are 29. In the studies with new generation DKs, non-ergonomic interactions had an increased presence of VRISE than the ones with ergonomic interactions. Also, in the studies which used DKs with non-ergonomic interactions, 7 participants out of 158 dropped out, while in studies with ergonomic interactions, only one participant out of 218 dropped out. Importantly, when new generation CVs with ergonomic interactions are exclusively considered, there are no VRISE or dropouts in any of the 11 studies with 546 participants. Finally, VRISE were only present in one study using a new generation CV HMD, where 3 participants dropped out. This study was the only one with a new generation CV HMD that did not involve ergonomic interactions.

The Bayesian independent samples *t*-test highlighted that studies involving new generation VR HMDs have significantly less frequent VRISE ($BF_{10} = 144.68; p < 0.001$). The difference in the existence of dropouts was not substantial, yet the studies with new generation HMDs have less frequent dropouts ($BF_{10} = 4.69; p < 0.05$) than studies with old HMDs. Notably, the studies which used a new generation CV HMD have significantly less frequent VRISE ($BF_{10} = 46.39; p < 0.001$) but not less frequent dropouts ($BF_{10} = 1.66; p = 0.16$) than the studies which used a new generation DK HMD. Finally, the studies which implemented VR software with ergonomic interactions had substantially less frequent VRISE ($BF_{10} = 19.54; p < 0.001$) and dropouts ($BF_{10} = 16.01; p < 0.001$) than studies which used VR software with non-ergonomic interactions.

The Bayesian Pearson's correlations demonstrated a substantial negative correlation between the presence of VRISE and the HMD generation [$BF_{10} = 328.03; r_{(44)} = -0.56, p < 0.001$], while the presence of VRISE robustly demonstrated a positive correlation with the existence of dropouts [$BF_{10} = 83510.53; r_{(44)} = 0.68, p < 0.001$]. Also, the utilization of ergonomic interactions was significantly negatively correlated with VRISE [$BF_{10} = 20.11; r_{(44)} = -0.42, p < 0.001$] and the existence of dropouts [$BF_{10} = 16.11; r_{(44)} = -0.41, p < 0.001$].

Discussion

The results of the meta-analysis indicated that VR HMDs have been implemented in diverse clinical conditions and age groups, as well as the unquestionable difference between old generation and new generation HMDs. There were significantly more frequent VRISE in the studies involving old generation VR HMDs compared to studies with new generation HMDs. Additionally, the frequency of VRISE correlated negatively with the HMD generation. Hence, the older the utilized HMD, the higher the VRISE frequency. Moreover, the existence of dropouts significantly and positively correlated with the presence of VRISE.

Nevertheless, one potential reason for the higher dropouts in old generation studies is that several studies included follow-up sessions (e.g., VRET) and participants may have opted not to return to complete the remaining sessions for reasons other than the presence of VRISE. However, in the old generation studies, the dropout rates were low in relation to the size of the population, albeit there were VRISE present. The low dropout rates in the old generation HMD studies may be due to the fixed intervals between the VR sessions where the participants were able to rest and obtain relief from any adverse effects they were experiencing.

Furthermore, the incidence of VRISE in old generation HMD studies may be due to anxiety levels (Bouchard et al., 2011) or be self-induced (Almeida et al., 2017) as several of the studies had either stress-related aims or included participants with stress disorders. However, several of the new generation HMD studies also had comparable aims and/or populations and included patients with clinical conditions (e.g., Alzheimer's disease, Parkinson's disease, stroke, and movement disorders) which have high comorbidity with stress and anxiety (Factor et al., 1995; Smith et al., 2000; Jenner, 2003; Allen and Bayraktutan, 2009). Also, the rates of self-induced VRISE are expected to be equal in both new and old generation HMD studies. In addition, the reporting of VRISE may be for reasons unrelated to the quality of the hardware or software (e.g., subjectivity in the reporting of VRISE, individual differences in the experience of VRISE) (Kortum and Peres, 2014; Almeida et al., 2017). However, this modulation is again expected to have affected both new and old generation HMD studies in a similar way.

Beyond the difference between old and new generations HMDs, a substantial difference is observed between DK and CV new generation HMDs. Significantly fewer VRISE were present in the studies that used a CV, indicating the superiority of new generation CV HMDs compared to new generation DK HMDs. Furthermore, the studies (i.e., both old and new generation studies) which utilized VR software with ergonomic interactions had robustly less frequent VRISE and dropouts than the studies which implemented VR software with non-ergonomic interactions. However, the ergonomic interactions do not appear to mitigate the dropout frequency and the incidence of VRISE in old generation HMDs. In contrast, VRISE were present in more DK studies with non-ergonomic interactions compared to DK studies with ergonomic interactions. Similarly, more participants dropped out from DK studies with non-ergonomic interactions. Notably, there were no VRISE or dropouts in CV studies with ergonomic interactions. Therefore, the contribution of ergonomic interactions in the reduction of VRISE increases when newer and better HMDs are utilized. To conclude, the findings of the meta-analysis are aligned with the outcomes of the technological review.

GENERAL DISCUSSION

Technological Competence in VR Neuroscience and Neuropsychology

The findings of our technological literature review suggest that the hardware features of old generation HMDs and new generation DKs do not meet the minimum hardware features that alleviate or eradicate VRISE. Instead, the technological literature review postulates the suitability of new generation CVs which have specific hardware capabilities to alleviate VRISE. However, VR software attributes (e.g., ergonomic interactions) are equally vital.

Secondly, the findings of our meta-analysis of 44 neuroscientific or neuropsychological studies using VR are aligned with the outcomes of our technological review, where VRISE were substantially less frequent in studies which utilized new generation VR HMDs. In particular, the studies which used new generation CVs accompanied by ergonomic interactions did not have any VRISE or dropouts. Therefore, the combined outcomes of the technological review and the meta-analysis indicate that the appropriate VR HMDs are those with hardware characteristics equal to or greater than the HTC Vive and Oculus Rift, though the VR HMD should be implemented in conjunction with VR software which offers ergonomic interactions.

However, researchers may have to opt for an HMD based on their available budget. For example, the Oculus Rift costs around $400, while the HTC Vive costs around $500. Moreover, the majority of HMDs also require a VR-ready desktop PC or a laptop to be operated, so a researcher needs to additionally spend around $500–$1,500 for a desktop computer or laptop to utilize these HMDs. Hence, the combined cost would be between $800 and $1,900. The cost of VR equipment (e.g., both HMD and computer) may lead researchers to use HMDs that are cheaper, albeit that they are more likely to result in VRISE. However, in the market, there are plenty of cost-effective alternatives that meet the minimum hardware criteria. For example, the Oculus Quest is a standalone HMD (i.e., it does not require a PC, a laptop, or a smartphone to be operated) and it costs approximately $400. Hence, a researcher can spend the equivalent of the price of a neuropsychological test or a smartphone to acquire and use an HMD that meets the minimum hardware criteria to lower the presence of VRISE.

Nonetheless, the selection of an appropriate VR HMD and software requires technological competency from the researchers, clinicians, and/or research software developers. Unfortunately, the meta-analysis results do not indicate that technological knowledge of VR has been well-established in neuroscience. Of course, the utilization of old generation HMDs and new generation DKs pre-2016 is justified as the new generation CVs were not available. However, in our meta-analysis, 25 studies were conducted between 2016 and 2019, where half of these studies (13/25) implemented an inappropriate HMD (i.e., old generation HMD or new generation DK). However, 10 studies used a DK2 which has a marginally lower resolution than the minimum hardware criteria, while our meta-analysis results indicated that its utilization in conjunction with ergonomic interactions appears to alleviate the frequency of VRISE, but not as effectively as the CV HMDs. Furthermore, one fifth of the studies did use a new generation HMD, but they did not have ergonomic interactions in their VR software. Therefore, at this time, VR technological competence does not seem to have been well-established in neuroscience. As a result, in the studies since 2016, the health and safety of the participants may not substantially guaranteed, and the reliability of the results may be questionable, as VRISE substantially decreases reaction times and overall cognitive performance (Plant and Turner, 2009; Nalivaiko et al., 2015; Plant, 2016; Nesbitt et al., 2017; Mittelstaedt et al., 2018), as well as confounding neuroimaging and physiological data (Toschi et al., 2017; Arafat et al., 2018; Gavgani et al., 2018). The selection of an appropriate HMD is paramount for successfully implementing VR HMDs in cognitive neuroscience and neuropsychology.

However, the implementation of the currently available and appropriate HMDs in neuroscience and neuropsychology should be compatible with the research aims. For example, in research designs where the user should be active (i.e., navigating, walking, and interacting within the VE) instead of being idle, or in a standing or a seated position, the researcher should opt for the best HMD that permits intense body movement and activity. In this setting, the Oculus Rift was found to be inferior to the HTC Vive on pick-and-place (i.e., relocating objects) tasks, whilst the HTC Vive also provided a substantially superior VR experience for users compared to the Oculus Rift (Suznjevic et al., 2017). Moreover, the HTC Vive provides an interactive area that is twice the size (25 m^2) of the Oculus Rift, albeit that both are very accurate in tracking (Borrego et al., 2018). Nevertheless, the HTC Vive was found to lose motion-tracking and the ground level becomes slanted when the user goes out of bounds (Niehorster et al., 2017). This shortcoming solely affects studies where the participant needs to go out of the tracking area. In most neuroscientific designs, the recommended maximum play area by HTC (6.25 m^2) or by Borrego et al. (2018) (25 m^2) are both substantially adequate for conducting ecological valid experiments (Borges et al., 2018; Borrego et al., 2018). Nonetheless, the slanted floor or lost tracking is not a hardware problem but a software one (Borges et al., 2018). In cases where the participant is required to go out of the tracking area, the tracking problem or the slanted floor may be easily corrected by adding 3 additional trackers (Peer et al., 2018), using software with an improved algorithm (freely distributed by NASA Ames Research Center) (Borges et al., 2018), or by simply updating the firmware of the lighthouse base stations. In summary, the researchers should be technologically competent to not only identify and implement a safe HMD and software, but an HMD and software that facilitate the optimal research methods pertinent to their research needs and aims.

As discussed in our technological review, the quality of the implemented VR software is equally important to avoid VRISE. Our meta-analysis of VR studies indicated that the utilization of ergonomic interactions is crucial albeit with the utilization of an appropriate HMD. For example, Detez et al. (2019) used the HTC Vive to investigate physiological arousal and behavior during gambling. However, the interactions and navigation within the VE were facilitated by using a typical controller (Detez et al., 2019). Hence, their VR software did not support the utilization of the ergonomic 6DoF controllers (both hands) of the HTC Vive, which facilitate naturalistic navigation (e.g., teleportation) and interaction within the VE. Consequently, Detez et al.'s (2019) participants experienced VRISE and 3 of their participants discontinued their sessions and so their data were discarded (Detez et al., 2019). Importantly, Detez et al. (2019) only reported the presence of VRISE and dropout size. They did not provide any quantitative data on the intensity of VRISE, or the quality of their software attributes (e.g., graphics, sound, tutorials, in-game instructions and prompts) (Detez et al., 2019). Indeed, only six of the studies in the meta-analysis provided adequately explicit reports on VRISE and VR software. Since Detez et al. (2019) assessed reaction times and heart rates, these data are likely to be affected by VRISE, despite the study having a rigorous experimental design and using the HTC Vive. Therefore, it is important to use appropriate VR software and external hardware to prevent risks to the health and safety of the participants as well as the reliability of the results.

Limitations and Future Studies

The above technological review and meta-analysis of VR studies evidenced the importance of technological and methodological features in VR research and clinical designs. However, our meta-analysis of VR studies has some limitations. The meta-analysis considered VR studies with diverse populations and designs, which may have affected the frequency of VRISE and the existence of dropouts. Uniformity across studies (e.g., considering only VRET, assessments or a specific clinical population) was not possible due to the scarcity of neuroscience studies involving VR, especially using new generation HMDs. Moreover, the review did not consider any software details due to the scarceness of such descriptions in published studies. Future VR studies should report software and hardware features to allow an in-depth meta-analysis. Equally, only six studies provided quantitative reports of VRISE intensity, consequently, only the presence or absence of VRISE was considered. The dichotomous consideration of VRISE is susceptible to reports based on subjective criteria and individual differences, but this is likely to have affected the VRISE rates in both old and new generation studies. Future studies should aim to appraise the quality of the software and intensity of VRISE (e.g., using questionnaires). Studies should also attempt to clarify the acceptable duration of immersive VR sessions, which will aid researchers in designing their studies appropriately. Importantly, the cost of the VR software development should also be considered. Finally, studies should attempt to provide software development guidelines that enable researchers and/or research software developers to develop VR research software without depending on third parties (e.g., freelance developers or software development companies) and these guidelines should embed suggestions and instructions for VR software development, which meet the criteria discussed above.

Conclusion

The use of VR HMDs is becoming more popular in neuroscience either for clinical or research purposes and VR technology and methods have been well-accepted by diverse populations in terms of age groups and clinical conditions. A more pleasant VR experience and a reduction in VRISE symptomatology has been found using new generation CV HMDs, which deliver an adequately high display resolution, rapid image refresh rate, ergonomic design and has controllers which allow naturalistic navigation and movement within the VE environment, especially when there is restricted teleportation. The outcomes of the current technological review and meta-analysis support the feasibility of new generation VR CV HMDs to be implemented in cognitive neuroscience and neuropsychology. The findings of the technological review suggest methods that should be considered in the development or selection of VR research software, as well as hardware and software features that should be included in the research protocol. The selected VR HMD and the VR research software should enable suitable ergonomic interactions, locomotion techniques (e.g., teleportation), and kinetic mechanics which ensure VRISE are reduced or completely avoided. A meticulous approach and technological competence are compulsory to consolidate the viability of VR research and clinical designs in cognitive neuroscience and neuropsychology.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript/supplementary files.

AUTHOR CONTRIBUTIONS

PK had the initial idea and contributed to every aspect of this study. SC, LD, and SM contributed to the methodological aspects and the discussion of the results.

REFERENCES

Allen, C. L., and Bayraktutan, U. (2009). Oxidative stress and its role in the pathogenesis of ischaemic stroke. *Int. J. Stroke* 4, 461–470. doi: 10.1111/j.1747-4949.2009.00387.x

Almeida, A., Rebelo, F., Noriega, P., and Vilar, E. (2017). "Virtual reality self induced cybersickness: an exploratory study," in *International Conference on Applied Human Factors and Ergonomics* (Cham: Springer), 26–33. doi: 10.1007/978-3-319-60582-1_3

Anthes, C., García-Hernández, R. J., Wiedemann, M., and Kranzlmüller, D. (2016). "State of the art of virtual reality technology," in *2016 IEEE Aerospace Conference* (Big Sky, MT), 1–19. doi: 10.1109/AERO.2016.7500674

Arafat, I. M., Ferdous, S. M. S., and Quarles, J. (2018). "Cybersickness-provoking virtual reality alters brain signals of persons with multiple sclerosis." in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Reutlingen: IEEE), 1–120. doi: 10.1109/VR.2018.8446194

Banakou, D., Kishore, S., and Slater, M. (2018). Virtually being Einstein results in an improvement in cognitive Task performance and a decrease in age bias. *Front. Psychol.* 9:917. doi: 10.3389/fpsyg.2018.00917

Banville, F., Nolin, P., Lalonde, S., Henry, M., Dery, M. P., and Villemure, R. (2010). Multitasking and prospective memory: can virtual reality be useful for diagnosis? *Behav. Neurol.* 23, 209–211. doi: 10.1155/2010/320707

Barberia, I., Oliva, R., Bourdin, P., and Slater, M. (2018). Virtual mortality and near-death experience after a prolonged exposure in a shared virtual reality may lead to positive life-attitude changes. *PLoS ONE* 13:e0203358. doi: 10.1371/journal.pone.0203358

Bioulac, S., Lallemand, S., Rizzo, A., Philip, P., Fabrigoule, C., and Bouvard, M. P. (2012). Impact of time on task on ADHD patient's performances in a virtual classroom. *Eur. J. Paediatr. Neurol.* 16, 514–521. doi: 10.1016/j.ejpn.2012.01.006

Bland, M. (2015). *An Introduction to Medical Statistics, 4th Edn.* Oxford: Oxford University Press.

Bohil, C. J., Alicea, B., and Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* 12:752. doi: 10.1038/nrn3122

Borges, M., Symington, A., Coltin, B., Smith, T., and Ventura, R. (2018). "HTC Vive: analysis and accuracy improvement," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Madrid: IEEE), 2610–2615. doi: 10.1109/IROS.2018.8593707

Borrego, A., Latorre, J., Alcaniz, M., and Llorens, R. (2018). Comparison of oculus rift and HTC vive: feasibility for virtual reality-based exploration, navigation, exergaming, and rehabilitation. *Games Health J.* 7, 151–156. doi: 10.1089/g4h.2017.0114

Botella, C., García-Palacios, A., Villa, H., Baños, R. M., Quero, S., Alcañiz, M., et al. (2007). Virtual reality exposure in the treatment of panic disorder and agoraphobia: a controlled study. *Clin. Psychol. Psychother.* 14, 164–175. doi: 10.1002/cpp.524

Bouchard, S., Robillard, G., Renaud, P., and Bernier, F. (2011). Exploring new dimensions in the assessment of virtual reality induced side effects. *J. Comput. Inform. Technol.* 1, 20–32. Available online at: http://pubs.drdc-rddc.gc.ca/BASIS/pcandid/www/engpub/DDW?W%3DSYSNUM=536530&r=0

Bourdin, P., Barberia, I., Oliva, R., and Slater, M. (2017). A virtual out-of-body experience reduces fear of death. *PLoS ONE* 12:e0169343. doi: 10.1371/journal.pone.0169343

Bozgeyikli, E., Raij, A., Katkoori, S., and Dubey, R. (2016). "Point and teleport locomotion technique for virtual reality," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (Austin, TX: ACM), 205–216. doi: 10.1145/2967934.2968105

Brade, J., Dudczig, M., and Klimant, P. (2018). "Using virtual prototyping technologies to evaluate human-machine-interaction concepts," in *AW&I Conference* (Chemnitz), 3.

Brennesholtz, M. S. (2018). "3-1: Invited Paper: VR standards and guidelines," in *SID Symposium Digest of Technical Papers, 49* (Pleasantville, NY), 1–4. doi: 10.1002/sdtp.12476

Caputo, F., Greco, A., Egidio, D. A., Notaro, I., and Spada, S. (2017). "A preventive ergonomic approach based on virtual and immersive reality," in *International Conference on Applied Human Factors and Ergonomics* (Cham: Springer), 3–15. doi: 10.1007/978-3-319-60582-1_1

Carlozzi, N. E., Gade, V., Rizzo, A. S., and Tulsky, D. S. (2013). Using virtual reality driving simulators in persons with spinal cord injury: three screen display versus head mounted display. *Disabil. Rehabil.* 8, 176–180. doi: 10.3109/17483107.2012.699990

Christou, C. G., Michael-Grigoriou, D., Sokratous, D., and Tsiakoulia, M. (2018). "A virtual reality loop and wire game for stroke rehabilitation," in *28th International Conference on Artificial Reality and 23rd Telexistence and Eurographics Symposium on Virtual Environments* (Limassol).

Collins, M. K., Ding, V. Y., Ball, R. L., Dolce, D. L., Henderson, J. M., and Halpern, C. H. (2018). Novel application of virtual reality in patient engagement for deep brain stimulation: a pilot study. *Brain Stimul.* 11, 935–937. doi: 10.1016/j.brs.2018.03.012

Costello, P. J. (1997). *Health and Safety Issues Associated With Virtual Reality: A Review of Current Literature.* Loughborough: Advisory Group on Computer Graphics.

Cox, D. R., and Donnelly, C. A. (2011). *Principles of Applied Statistics.* Cambridge: Cambridge University Press

Davis, S., Nesbitt, K., and Nalivaiko, E. (2015). "Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters," in *Proceedings of the 11th Australasian Conference on Interactive Entertainment (IE 2015), 27* (Sydney), 30.

de França, A. C. P., and Soares, M. M. (2017). "Review of virtual reality technology: an ergonomic approach and current challenges," in *International Conference on Applied Human Factors and Ergonomics* (Cham: Springer), 52–61. doi: 10.1007/978-3-319-60582-1_6

Detez, L., Greenwood, L. M., Segrave, R., Wilson, E., Chandler, T., Ries, T., et al. (2019). A psychophysiological and behavioural study of slot machine near-misses using immersive virtual reality. *J. Gambl. Stud.* 35, 929–944. doi: 10.1007/s10899-018-09822-z

Factor, S. A., Podskalny, G. D., and Molho, E. S. (1995). Psychogenic movement disorders: frequency, clinical profile, and characteristics. *J. Neurol. Neurosurg. Psychiatry* 59, 406–412. doi: 10.1136/jnnp.59.4.406

Figueiredo, L., Rodrigues, E., Teixeira, J., and Techrieb, V. (2018). A comparative evaluation of direct hand and wand interactions on consumer devices. *Comput. Graph.* 77, 108–121. doi: 10.1016/j.cag.2018.10.006

Foerster, R. M., Poth, C. H., Behler, C., Botsch, M., and Schneider, W. X. (2016). Using the virtual reality device Oculus Rift for neuropsychological assessment of visual processing capabilities. *Sci. Rep.* 6:37016. doi: 10.1038/srep37016

Freeman, D., Bradley, J., Antley, A., Bourke, E., DeWeever, N., Evans, N., et al. (2016). Virtual reality in the treatment of persecutory delusions: randomised controlled experimental study testing how to reduce delusional conviction. *Br. J. Psychiatry* 209, 62–67. doi: 10.1192/bjp.bp.115.176438

Freeman, D., Evans, N., Lister, R., Antley, A., Dunn, G., and Slater, M. (2014). Height, social comparison, and paranoia: an immersive virtual reality experimental study. *Psychiatry Res.* 218, 348–352. doi: 10.1016/j.psychres.2013.12.014

Frommel, J., Sonntag, S., and Weber, M. (2017). "Effects of controller-based locomotion on player experience in a virtual reality exploration game," in *Proceedings of the 12th International Conference on the Foundations of Digital Games* (Hyannis, MA: ACM), 30. doi: 10.1145/3102071.3102082

Gaggioli, A., Pallavicini, F., Morganti, L., Serino, S., Scaratti, C., Briguglio, M., et al. (2014). Experiential virtual scenarios with real-time monitoring (interreality) for the management of psychological stress: a block randomized controlled trial. *J. Med. Internet Res.* 16:e167. doi: 10.2196/jmir.3235

Gavgani, A. M., Wong, R. H., Howe, P. R., Hodgson, D. M., Walker, F. R., and Nalivaiko, E. (2018). Cybersickness-related changes in brain hemodynamics: a pilot study comparing transcranial Doppler and near-infrared spectroscopy assessments during a virtual ride on a roller coaster. *Physiol. Behav.* 191, 56–64. doi: 10.1016/j.physbeh.2018.04.007

George, C., Demmler, M., and Hussmann, H. (2018). "Intelligent interruptions for IVR: investigating the interplay between presence, workload and attention," in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, QC: ACM), LBW511. doi: 10.1145/3170427.3188686

Gómez-Jordana, L. I., Stafford, J., Peper, C. L. E., and Craig, C. M. (2018). Virtual footprints can improve walking performance in people with Parkinson's disease. *Front. Neurol.* 9:681. doi: 10.3389/fneur.2018.00681

Goradia, I., Doshi, J., and Kurup, L. (2014). A review paper on oculus rift and project morpheus. *Int. J. Curr. Eng. Technol.* 4, 3196–3200. Available online at: http://inpressco.com/wp-content/uploads/2014/09/Paper183196-3200.pdf

Gromala, D., Tong, X., Shaw, C., Amin, A., Ulas, S., and Ramsay, G. (2016). Mobius floe: an immersive virtual reality game for pain distraction. *Electron. Imaging* 2016, 1–5. doi: 10.2352/ISSN.2470-1173.2016.4.ERVR-413

Hartanto, D., Kampmann, I. L., Morina, N., Emmelkamp, P. G., Neerincx, M. A., and Brinkman, W. P. (2014). Controlling social stress in virtual reality environments. *PLoS ONE* 9:e92804. doi: 10.1371/journal.pone.0092804

Hasler, B. S., Spanlang, B., and Slater, M. (2017). Virtual race transformation reverses racial in-group bias. *PLoS ONE* 12:e0174965. doi: 10.1371/journal.pone.0174965

Hecht, J. (2016). Optical dreams, virtual reality. *Optics Photon. News* 27, 24–31. doi: 10.1364/OPN.27.6.000024

Held, L., and Ott, M. (2018). On p-values and Bayes factors. *Annu. Rev. Stat. Appl.* 5, 393–419. doi: 10.1146/annurev-statistics-031017-100307

Hsieh, T. J. T., Kuo, Y. H., and Niu, C. K. (2018). "Utilizing HMD VR to improve the spatial learning and wayfinding effects in the virtual maze," in *International Conference on Human-Computer Interaction* (Cham: Springer), 38–42. doi: 10.1007/978-3-319-92285-0_6

JASP Team (2018). *JASP (Version 0.9) [Computer Software].*

Jenner, P. (2003). Oxidative stress in Parkinson's disease. *Ann. Neurol.* 53, S26–S38. doi: 10.1002/ana.10483

Jerald, J., LaViola, J. J. Jr., and Marks, R. (2017). "VR interactions," in *ACM SIGGRAPH 2017 Courses* (Los Angeles, CA: ACM), 19. doi: 10.1145/3084873.3084900

Kane, R. L., and Parsons, T. D. (eds.). (2017). *The Role of Technology in Clinical Neuropsychology.* New York, NY: Oxford University Press.

Kelly, J. W., Cherep, L. A., and Siegel, Z. D. (2017). Perceived space in the HTC vive. *ACM Trans. Appl. Percept.* 15:2. doi: 10.1145/3106155

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviation Psychol.* 3, 203–220. doi: 10.1207/s15327108ijap0303_3

Kim, A., Darakjian, N., and Finley, J. M. (2017). Walking in fully immersive virtual environments: an evaluation of potential adverse effects in older adults and individuals with Parkinson's disease. *J. Neuroeng. Rehabil.* 14:16. doi: 10.1186/s12984-017-0225-2

Kim, H. K., Park, J., Choi, Y., and Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. *Appl. Ergon.* 69, 66–73. doi: 10.1016/j.apergo.2017.12.016

Kim, J. W., Choe, W. J., Hwang, K. H., and Kwag, J. O. (2017). "78-2: the optimum display for virtual reality," in *SID Symposium Digest of Technical Papers,* 48, 1146–1149. doi: 10.1002/sdtp.11845

Kim, K., Kim, J., Ku, J., Kim, D. Y., Chang, W. H., Shin, D. I., et al. (2004). A virtual reality assessment and training system for unilateral neglect. *Cyberpsychol. Behav.* 7, 742–749. doi: 10.1089/cpb.2004.7.742

Kortum, P., and Peres, S. C. (2014). The relationship between system effectiveness and subjective usability scores using the System Usability Scale. *Int. J. Human Comput. Int.* 30, 575–584. doi: 10.1080/10447318.2014.904177

Lubetzky, A. V., Kary, E. E., Harel, D., Hujsak, B., and Perlin, K. (2018). Feasibility and reliability of a virtual reality oculus platform to measure sensory integration for postural control in young adults. *Physiother. Theory Pract.* 34, 935–950. doi: 10.1080/09593985.2018.1431344

Marsman, M., and Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *Eur. J. Dev. Psychol.* 14, 545–555. doi: 10.1080/17405629.2016.1259614

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

Meyerbroeker, K., Morina, N., Kerkhof, G. A., and Emmelkamp, P. M. G. (2013). Virtual reality exposure therapy does not provide any additional value in agoraphobic patients: a randomized controlled trial. *Psychother. Psychosom.* 82, 170–176. doi: 10.1159/000342715

Mittelstaedt, J. M., Wacker, J., and Stelling, D. (2018). VR aftereffect and the relation of cybersickness and cognitive performance. *Virtual Real.* 23, 143–154. doi: 10.1007/s10055-018-0370-3

Montenegro, J. M. F., and Argyriou, V. (2017). Cognitive evaluation for the diagnosis of Alzheimer's disease based on turing test and virtual environments. *Physiol. Behav.* 173, 42–51. doi: 10.1016/j.physbeh.2017.01.034

Moreau, G., Guay, M. C., Achim, A., Rizzo, A., and Lageix, P. (2006). The virtual classroom: an ecological version of the continuous performance test–A pilot study. *Annu. Rev. CyberTherapy Telemed.* 4, 59–66. doi: 10.1037/e695432011-008

Mottelson, A., and Hornbæk, K. (2017). "Virtual reality studies outside the laboratory," in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (Gothenburg: ACM), 9. doi: 10.1145/3139131.3139141

Nalivaiko, E., Davis, S. L., Blackmore, K. L., Vakulin, A., and Nesbitt, K. V. (2015). Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiol. Behav.* 151, 583–590. doi: 10.1016/j.physbeh.2015.08.043

Nesbitt, K., Davis, S., Blackmore, K., and Nalivaiko, E. (2017). Correlating reaction time and nausea measures with traditional measures of cybersickness. *Displays* 48, 1–8. doi: 10.1016/j.displa.2017.01.002

Niehorster, D. C., Li, L., and Lappe, M. (2017). The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research. *i-Perception* 8:2041669517708205. doi: 10.1177/2041669517708205

Oagaz, H., Schoun, B., Pooji, M., and Choi, M. H. (2018). Neurocognitive assessment in virtual reality through behavioral response analysis. *IEEE J. Biomed. Health Inform.* 23, 1899–1910. doi: 10.1109/JBHI.2018.2881455

Palmisano, S., Mursic, R., and Kim, J. (2017). Vection and cybersickness generated by head-and-display motion in the Oculus Rift. *Displays* 46, 1–8. doi: 10.1016/j.displa.2016.11.001

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front. Human Neurosci.* 9:660. doi: 10.3389/fnhum.2015.00660

Parsons, T. D., and Barnett, M. (2017). Virtual Apartment-Based Stroop for assessing distractor inhibition in healthy aging. *Appl. Neuropsychol. Adult* 26, 144–154.doi: 10.1080/23279095.2017.1373281

Parsons, T. D., and Carlew, A. R. (2016). Bimodal virtual reality stroop for assessing distractor inhibition in autism spectrum disorders. *J. Autism Dev. Disord.* 46, 1255–1267. doi: 10.1007/s10803-015-2663-7

Parsons, T. D., Courtney, C. G., and Dawson, M. E. (2013). Virtual reality Stroop task for assessment of supervisory attentional processing. *J. Clin. Exp. Neuropsychol.* 35, 812–826. doi: 10.1080/13803395.2013.824556

Parsons, T. D., and McMahan, T. (2017). An initial validation of the virtual of the virtual environment grocery store. *J. Neurosci. Methods* 291, 13–19. doi: 10.1016/j.jneumeth.2017.07.027

Parsons, T. D., McMahan, T., and Kane, R. (2018). Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms. *Clin. Neuropsychol.* 32, 16–41. doi: 10.1080/13854046.2017.1337932

Parsons, T. D., Rizzo, A. A., Bamattre, J., and Brennan, J. (2007). Virtual reality cognitive performance assessment test. *Annu. Rev. CyberTherapy Telemed.* 5, 163–171. doi: 10.1037/e688982007-001

Peck, T. C., Seinfeld, S., Aglioti, S. M., and Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Conscious. Cogn.* 22, 779–787. doi: 10.1016/j.concog.2013.04.016

Peer, A., Ullich, P., and Ponto, K. (2018). "Vive tracking alignment and correction made easy," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Reutlingen: IEEE), 653–654. doi: 10.1109/VR.2018.8446435

Plant, R. R. (2016). A reminder on millisecond timing accuracy and potential replication failure in computer-based psychology experiments: an open letter. *Behav. Res. Methods* 48, 408–411. doi: 10.3758/s13428-015-0577-0

Plant, R. R., and Turner, G. (2009). Millisecond precision psychological research in a world of commodity computers: new hardware, new problems? *Behav. Res. Methods* 41, 598–614. doi: 10.3758/BRM.41.3.598

Plouzeau, J., Paillot, D., Chardonnet, J. R., and Merienne, F. (2015). "Effect of proprioceptive vibrations on simulator sickness during navigation task in virtual environment," in *International Conference on Artificial Reality and Telexistence* (Kyoto).

Porcino, T. M., Clua, E., Trevisan, D., Vasconcelos, C. N., and Valente, L. (2017). "Minimizing cyber sickness in head mounted display systems: design guidelines and applications," in *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)* (Perth, WA: IEEE), 1–6. doi: 10.1109/SeGAH.2017.7939283

Quinlivan, B., Butler, J. S., Beiser, I., Williams, L., McGovern, E., O'Riordan, S., et al. (2016). Application of virtual reality head mounted display for investigation of movement: a novel effect of orientation of attention. *J. Neural Eng.* 13:056006. doi: 10.1088/1741-2560/13/5/056006

Rakkolainen, I., Turk, M., and Höllerer, T. (2016). "A superwide-FOV optical design for head-mounted displays," in *Proceedings of the Joint 26th International Conference on Artificial Reality and Telexistence and the 21th Eurographics Symp. on Virtual Environments (ICAT-EGVE 2016)* (Little Rock, AR), 45–48.

Reger, G. M., Holloway, K. M., Candy, C., Rothbaum, B. O., Difede, J., Rizzo, A. A., et al. (2011). Effectiveness of virtual reality exposure therapy for active duty soldiers in a military mental health clinic. *J. Traumat. Stress* 24, 93–96. doi: 10.1002/jts.20574

Rizzo, A., Difede, J., Rothbaum, B. O., Reger, G., Spitalnick, J., Cukor, J., et al. (2010). Development and early evaluation of the Virtual Iraq/Afghanistan exposure therapy system for combat-related PTSD. *Ann. N. Y. Acad. Sci.* 1208, 114–125. doi: 10.1111/j.1749-6632.2010.05755.x

Rizzo, A. A., Schultheis, M., Kerns, K. A., and Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychol. Rehabil.* 14, 207–239. doi: 10.1080/09602010343000183

Rooney, B., Burke, C., Bálint, K., O'Leary, T., Parsons, T., Lee, C. T., et al. (2017). "Virtual reality, presence and social cognition: the effect of eye-gaze and narrativity on character engagement," in *2017 23rd International Conference on Virtual System and Multimedia (VSMM)* (Dublin: IEEE), 1–6. doi: 10.1109/VSMM.2017.8346272

Rothbaum, B. O., Price, M., Jovanovic, T., Norrholm, S. D., Gerardi, M., Dunlop, B., et al. (2014). A randomized, double-blind evaluation of D-cycloserine or alprazolam combined with virtual reality exposure therapy for posttraumatic stress disorder in Iraq and Afghanistan War veterans. *Am. J. Psychiatry* 171, 640–648. doi: 10.1176/appi.ajp.2014.13121625

Rouder, J. N., and Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behav. Res.* 47, 877–903. doi: 10.1080/00273171.2012.734737

Sharples, S., Cobb, S., Moody, A., and Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): comparison of head mounted display (HMD), desktop and projection display systems. *Displays* 29, 58–69. doi: 10.1016/j.displa.2007.09.005

Shiban, Y., Schelhorn, I., Pauli, P., and Mühlberger, A. (2015). Effect of combined multiple contexts and multiple stimuli exposure in spider phobia: a randomized clinical trial in virtual reality. *Behav. Res. Therapy* 71, 45–53. doi: 10.1016/j.brat.2015.05.014

Slater, M., and Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. *Presence* 6, 603–616. doi: 10.1162/pres.1997.6.6.603

Smith, M. A., Rottkamp, C. A., Nunomura, A., Raina, A. K., and Perry, G. (2000). Oxidative stress in Alzheimer's disease. *Biochim. Biophys. Acta* 1502, 139–144. doi: 10.1016/S0925-4439(00)00040-5

Sportillo, D., Paljic, A., Boukhris, M., Fuchs, P., Ojeda, L., and Roussarie, V. (2017). "An immersive Virtual Reality system for semi-autonomous driving simulation: a comparison between realistic and 6-DoF controller-based interaction," in *Proceedings of the 9th International Conference on Computer and Automation Engineering* (Sydney: ACM), 6–10. doi: 10.1145/3057039.3057079

Stanney, K. M., and Hale, K. S. (2014). *Handbook of Virtual Environments: Design, Implementation, and Applications.* Boca Raton, FL: CRC Press. doi: 10.1201/b17360

Suznjevic, M., Mandurov, M., and Matijasevic, M. (2017). "Performance and QoE assessment of HTC vive and oculus rift for pick-and-place tasks in VR," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)* (Erfurt: IEEE), 1–3. doi: 10.1109/QoMEX.2017.7965679

Toschi, N., Kim, J., Sclocco, R., Duggento, A., Barbieri, R., Kuo, B., et al. (2017). Motion sickness increases functional connectivity between visual motion and nausea-associated brain regions. *Auton. Neurosci.* 202, 108–113. doi: 10.1016/j.autneu.2016.10.003

Veling, W., Brinkman, W. P., Dorrestijn, E., and Van Der Gaag, M. (2014). Virtual reality experiments linking social environment and psychosis: a pilot study. *Cyberpsychol. Behav. Social Netw.* 17, 191–195. doi: 10.1089/cyber.2012.0497

Viirre, E., Price, B. J., and Chase, B. (2014). "Direct effects of virtual environments on users," in *Handbook of Virtual Environments*, eds K. S. Hale and K. M. Stanney (Boca Raton, FL: CRC Press), 527–535. doi: 10.1201/b17360-28

Vorländer, M., and Shinn-Cunningham, B. (2014). "Virtual auditory displays," in *Handbook of Virtual Environments*, eds K. S. Hale and K. M. Stanney (Boca Raton, FL: CRC Press), 107–134. doi: 10.1201/b17360-7

Wetzels, R., and Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychon. Bull. Rev.* 19, 1057–1064. doi: 10.3758/s13423-012-0295-x

Yeh, S. C., Li, Y. Y., Zhou, C., Chiu, P. H., and Chen, J. W. (2018). Effects of virtual reality and augmented reality on induced anxiety. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 1345–1352. doi: 10.1109/TNSRE.2018.2844083

Zimmer, P., Buttlar, B., Halbeisen, G., Walther, E., and Domes, G. (2018). Virtually stressed? A refined virtual reality adaptation of the Trier Social Stress Test (TSST) induces robust endocrine responses. *Psychoneuroendocrinology.* 101, 186–192. doi: 10.1016/j.psyneuen.2018.11.010

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

In this chapter, the minimum hardware characteristics of immersive VR HMDs for the mitigation or avoidance of VRISE were identified. Equally, the required software features (e.g., ergonomic interactions) which further mitigate the frequency and intensity of VRISE were also identified and discussed. The meta-analysis of the selected neuroscience studies confirmed the importance of the identified hardware and software characteristics. Therefore, regarding the development of VR-EAL, the findings of this chapter allow the selection of new generation of HMDs (e.g., HTC Vive) which the VR-EAL should be compatible with, as well as the software features that the VR-EAL should incorporate. However, the meta-analysis of the studies also revealed that the researchers do not report the quality of the software and the intensity of VRISE in a quantitative way, which postulates the requirement for a tool (e.g., a questionnaire) which may measure both of them.

Chapter 3: Validation of the Virtual Reality Neuroscience Questionnaire: Maximum Duration of Immersive Virtual Reality Sessions Without the Presence of Pertinent Adverse Symptomatology

Chapter 2 highlighted the necessity for a tool capable of measuring both the quality of software features and the intensity of VRISE. Chapter 3 presents and discusses the development and validation of the virtual reality neuroscience questionnaire (VRNQ). Regarding the implementation of VR HMDs for research or clinical purposes, Chapter 3 also investigates the possible maximum duration of the VR sessions (i.e., the maximum duration of being immersed uninterruptedly), since there are not any guidelines for the possible maximum duration of VR sessions without experiencing intense VRISE. This study may offer a tool (i.e., VRNQ) which may be implemented for evaluating VR-EAL during its development, as well as may indicate the maximum duration that VR-EAL should have.

This chapter has been published in the Frontiers in Human Neuroscience: Kourtesis, P., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2019). Validation of the virtual reality neuroscience questionnaire: maximum duration of immersive virtual reality sessions without the presence of pertinent adverse symptomatology. Frontiers in Human Neuroscience, 13, 417. doi: https://doi.org/10.3389/fnhum.2019.00417

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 26 November 2019
doi: 10.3389/fnhum.2019.00417

Validation of the Virtual Reality Neuroscience Questionnaire: Maximum Duration of Immersive Virtual Reality Sessions Without the Presence of Pertinent Adverse Symptomatology

***Panagiotis Kourtesis*[1,2,3,4]*, *Simona Collina*[3,4], *Leonidas A. A. Doumas*[2] and *Sarah E. MacPherson*[1,2]**

[1] Human Cognitive Neuroscience, Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom, [2] Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom, [3] Lab of Experimental Psychology, Suor Orsola Benincasa University of Naples, Naples, Italy, [4] Interdepartmental Centre for Planning and Research "Scienza Nuova", Suor Orsola Benincasa University of Naples, Naples, Italy

OPEN ACCESS

Edited by:
Valerio Rizzo,
University of Palermo, Italy

Reviewed by:
Eugene Nalivaiko,
University of Newcastle, Australia
Mark Dennison,
United States Army Research Laboratory, United States
Justin Maximilian Mittelstädt,
Institute of Aerospace Medicine, German Aerospace Center (DLR), Germany

****Correspondence:***
Panagiotis Kourtesis
pkourtes@exseed.ed.ac.uk

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *12 August 2019*
Accepted: *11 November 2019*
Published: *26 November 2019*

Citation:
Kourtesis P, Collina S, Doumas LAA and MacPherson SE (2019) Validation of the Virtual Reality Neuroscience Questionnaire: Maximum Duration of Immersive Virtual Reality Sessions Without the Presence of Pertinent Adverse Symptomatology. Front. Hum. Neurosci. 13:417. doi: 10.3389/fnhum.2019.00417

There are major concerns about the suitability of immersive virtual reality (VR) systems (i.e., head-mounted display; HMD) to be implemented in research and clinical settings, because of the presence of nausea, dizziness, disorientation, fatigue, and instability (i.e., VR induced symptoms and effects; VRISE). Research suggests that the duration of a VR session modulates the presence and intensity of VRISE, but there are no suggestions regarding the appropriate maximum duration of VR sessions. The implementation of high-end VR HMDs in conjunction with ergonomic VR software seems to mitigate the presence of VRISE substantially. However, a brief tool does not currently exist to appraise and report both the quality of software features and VRISE intensity quantitatively. The Virtual Reality Neuroscience Questionnaire (VRNQ) was developed to assess the quality of VR software in terms of user experience, game mechanics, in-game assistance, and VRISE. Forty participants aged between 28 and 43 years were recruited (18 gamers and 22 non-gamers) for the study. They participated in 3 different VR sessions until they felt weary or discomfort and subsequently filled in the VRNQ. Our results demonstrated that VRNQ is a valid tool for assessing VR software as it has good convergent, discriminant, and construct validity. The maximum duration of VR sessions should be between 55 and 70 min when the VR software meets or exceeds the parsimonious cut-offs of the VRNQ and the users are familiarized with the VR system. Also, the gaming experience does not seem to affect how long VR sessions should last. Also, while the quality of VR software substantially modulates the maximum duration of VR sessions, age and education do not. Finally, deeper immersion, better quality of graphics and sound, and more helpful in-game instructions and prompts were found to reduce VRISE intensity. The VRNQ facilitates the brief assessment and reporting of the quality of VR software features and/or the intensity of VRISE, while its minimum and parsimonious cut-offs may appraise the suitability of VR software for implementation in research and clinical settings. The findings of this study contribute to the establishment of rigorous VR methods that are crucial for the viability of immersive VR as a research and clinical tool in cognitive neuroscience and neuropsychology.

Keywords: virtual reality, VRISE, VR sickness, cybersickness, neuroscience, neuropsychology, psychology, motion sickness

INTRODUCTION

Immersive virtual reality (VR) has emerged as a novel tool for neuroscientific and neuropsychological research (Bohil et al., 2011; Parsons, 2015; Parsons et al., 2018). Nevertheless, there are concerns pertinent to implementing VR in research and clinical settings, especially regarding the head-mounted display (HMD) systems (Sharples et al., 2008; Bohil et al., 2011; de França and Soares, 2017; Palmisano et al., 2017). A primary concern is the presence of adverse physiological symptoms (i.e., nausea, dizziness, disorientation, fatigue, and postural instability), which are referred to as motion-sickness, cybersickness, VR sickness or VR induced symptoms and effects (VRISE) (Sharples et al., 2008; Bohil et al., 2011; de França and Soares, 2017; Palmisano et al., 2017).

Longer durations in a virtual environment have been associated with a higher probability of experiencing VRISE, while the intensity of VRISE also appears to increase proportionally with the duration of the VR session (Sharples et al., 2008). However, extensive linear and angular accelerations provoke intense VRISE, even in a short period of time (McCauley and Sharkey, 1992; LaViola, 2000; Gavgani et al., 2018). VRISE may place the health and safety of the participants or patients at risk of experiencing adverse physiological symptoms (Parsons et al., 2018). Research has also shown that VRISE induce significant decreases in reaction times and overall cognitive performance (Nalivaiko et al., 2015; Nesbitt et al., 2017; Mittelstaedt et al., 2019), as well as substantially increasing body temperatures and heart rates (Nalivaiko et al., 2015), which may compromise physiological data acquisition. Furthermore, the presence of VRISE has been found to significantly augment cerebral blood flow and oxyhemoglobin concentration (Gavgani et al., 2018), electrical brain activity (Arafat et al., 2018), and the connectivity between stimulus-response regions and nausea-processing regions (Toschi et al., 2017). Thus, VRISE appear to confound the reliability of neuropsychological, physiological, and neuroimaging data (Kourtesis et al., 2019).

To our knowledge, there do not appear to be any guidelines as to the appropriate maximum duration of VR research and clinical sessions to evade or alleviate the presence of VRISE. Recently, our work has suggested that VRISE are substantially reduced or prevented by VR software that facilitates ergonomic navigation (e.g., physical movement) and interaction (e.g., direct-hand tracking) facilitated by the hardware capabilities (e.g., motion tracking) of commercial, contemporary VR HMDs comparable to or more advanced than the HTC Vive and/or Oculus Rift (Kourtesis et al., 2019). However, there are other factors such as the type of display and its features that may also induce or reduce VRISE (Mittelstaedt et al., 2018; Kourtesis et al., 2019). Nevertheless, we note that adequate technological competence is required to be able to implement appropriate VR hardware and/or software. In an attempt to reach a methodological consensus, we have proposed minimum hardware and software features, which appraise the suitability of VR hardware and software (see **Table 1**; Kourtesis et al., 2019).

While VRISE may occur for various reasons, they are predominantly the undesirable outcomes of hardware and software insufficiencies (e.g., low resolution and refresh rates of the image, a narrow field of view, non-ergonomic interactions, and inappropriate navigation modes) (de França and Soares, 2017; Palmisano et al., 2017; Kourtesis et al., 2019). In terms of hardware, the technical specifications of the computer (e.g., processing power and graphics card), and VR HMD (e.g., the field of view, refresh rate, and resolution) suffice to appraise their suitability (Kourtesis et al., 2019). However, there is not a tool to quantify the software's recommended features, as well as the intensity of VRISE (Kourtesis et al., 2019). Currently, the most frequently used measure of VRISE is the simulator sickness questionnaire (SSQ), which only considers the symptoms pertinent to simulator sickness (Kennedy et al., 1993). However, the SSQ does not assess software attributes (Kennedy et al., 1993), and there is an argument that simulator sickness symptomatology may not be identical to VRISE (Stanney et al., 1997). There is thus a need for a tool, which will enable researchers to assess both the suitability of VR software, as well as the intensity of VRISE.

Our recent technological literature review of VR hardware and software pinpointed four domains that should be considered in the development or selection of VR research/clinical software (Kourtesis et al., 2019). The domains are user experience, game mechanics, in-game assistance, and VRISE. Each domain has five criteria that should be met to ensure the appropriateness of the software (see **Table 1**). Also, in the same study, the meta-analysis of 44 VR neuroscientific studies revealed that most of the studies did not report quantitatively VR software's quality and/or VRISE intensity (Kourtesis et al., 2019). In an attempt to provide a brief tool for the appraisal of VR research/clinical software features and VRISE intensity, we developed the virtual reality neuroscience questionnaire (VRNQ), which includes twenty questions that address five criteria under each domain. This study aimed to validate the VRNQ and provide suggestions for the duration of VR research/clinical sessions. We also considered the gaming experience of the participants to examine whether this may affect the duration of the VR sessions. Lastly, we investigated the software predictors of VRISE as measured by the VRNQ.

MATERIALS AND METHODS

Participants

Forty participants (21 males) aged between 28 and 43 years ($M = 32.08$; $SD = 3.54$) and an educational level between 12 and 16 full-time years of education ($M = 14.25$; $SD = 1.37$) were recruited for the study. Eighteen participants (10 males) identified themselves as gamers through self-report and 22 as non-gamers (11 males). The gamer experience was a dichotomous variable (i.e., gamer or non-gamer) based on the participants' response to a question asking whether they played games on a weekly basis. The participants responded to a call disseminated through mailing lists at the University of Edinburgh and social media. The study was approved by the Philosophy, Psychology and Language Sciences Research Ethics Committee of the University of Edinburgh. All participants provided written informed consent prior to taking part.

Material

Hardware

An HTC Vive HMD with two lighthouse-stations for motion tracking was used with two HTC Vive's wands with 6 degrees of freedom (DoF) to facilitate navigation and interactions within the environment (Kourtesis et al., 2019). The VR area where the participants were immersed and interacted with the virtual environments was 4.4 m^2. Additionally, the HMD was connected to a laptop with an Intel Core i7 7700HQ processor at 2.80 GHz, 16 GB RAM, a 4095 MB NVIDIA GeForce GTX 1070 graphics card, a 931 GB TOSHIBA MQ01ABD100 (SATA) hard disk, and Realtek High Definition Audio.

Software

Three VR games were selected, which included ergonomic navigation (i.e., teleportation and physical mobility) and interactions (i.e., 6 DoF wands simulating hand movements) with the virtual environment. In line with Kourtesis et al. (2019), the VR software inclusion criteria (see **Table 1**) were: (1) ergonomic interactions which simulate real-life hand movements; (2) a navigation system which uses teleportation and physical mobility; (3) comprehensible tutorials pertinent to the controls; and (4) in-game instructions and prompts which assist the user in orientating and interacting with the virtual environment. The suitability of the VR software for both gamers and non-gamers was also considered. The selected VR games which met the above software criteria were: (1) "Job Simulator" (Session 1)[1]; (2) "The Lab" (Session 2)[2]; and (3) "Rick and Morty: Virtual Rick-ality" (Session 3)[3]. In "Job Simulator," the participant becomes an employee who has several occupations, such as a cook (preparing simply recipes), car mechanic (doing rudimentary tasks e.g., replacing faulty parts), and an office worker (making calls and sending emails). In "The Lab," the participant needs to complete several mini-games like slingshot (shooting down piles of boxes), longbow (shooting down invaders), xortex (spaceship-battles), postcards (visiting exotic places), human medical scan (exploring the human body), solar system (exploring the solar system), robot repair (repairing a robot), and secret shop (exploring a magical shop). In "Rick and Morty: Virtual Rick-ality," the participant needs to complete several imaginary home-chores as in "Job Simulator," though, in this case, the participant is required to follow a sequence of tasks according to a fictional storyline.

Virtual Reality Neuroscience Questionnaire (VRNQ)

The VRNQ measures the quality of user experience, game mechanics, and in-game assistance, as well as the intensity of VRISE. The VRNQ involves 20 questions where each question corresponds to one of the criteria for appropriate VR research/clinical software (e.g., the level of immersion; see **Table 1**). The 20 questions are grouped under four domains, where each domain encompasses five questions. Hence, VNRQ produces a total score corresponding to the overall quality of VR software, as well as four sub-scores (i.e., user experience, game mechanics, in-game assistance, VRISE). The user experience score is based on the intensity of the immersion, the level of enjoyment, as well as the quality of the graphics, sound, and VR technology (i.e., internal and external hardware). The game mechanics' score depends on the ease to navigate, physically move, and interact with the virtual environment (i.e., use, pick

TABLE 1 | Domains and criteria for VR research/clinical software.

Domains	User experience	Game mechanics	In-game assistance	VRISE
CRITERIA	An adequate level of immersion	A suitable navigation system (e.g., teleportation)	Digestible tutorials	Absence or insignificant presence of nausea
	Pleasant VR experience	Availability of physical movement	Helpful tutorials	Absence or insignificant presence of disorientation
	High quality graphics	Naturalistic picking/placing of items	Adequate duration of tutorials	Absence or insignificant presence of dizziness
	High quality sounds	Naturalistic use of items	Helpful in-game instructions	Absence or insignificant presence of fatigue
	Suitable hardware (HMD and computer)	Naturalistic 2-handed interaction	Helpful in-game prompts	Absence or insignificant presence of instability

Derived from Kourtesis et al. (2019).

---

[1] https://store.steampowered.com/app/448280/Job_Simulator/

[2] https://store.steampowered.com/app/450390/The_Lab/

[3] https://store.steampowered.com/app/469610/Rick_and_Morty_Virtual_Rickality/

and place, and hold items; two-handed interactions). The in-game assistance score appraises the quality of the tutorial(s), in-game instructions (e.g., description of the aim of the task), and prompts (e.g., arrows showing the direction). The VRISE are evaluated by the intensity of primary adverse symptoms and effects pertinent to VR (i.e., nausea, disorientation, dizziness, fatigue, and instability). VRNQ responses are indicated on a 7-point Likert style scale, ranging from 1 = extremely low to 7 = extremely high. The higher scores indicate a more positive outcome; this also applies to the evaluation of VRISE intensity. Hence, the higher VRISE score indicates a lower intensity of VRISE (i.e., 1 = extremely intense feeling, 2 = very intense feeling, 3 = intense feeling, 4 = moderate feeling, 5 = mild feeling, 6 = very mild feeling, 7 = absent). The VRNQ also includes space under each question, where the participant may provide optional qualitative feedback. For further details, please see the VRNQ in **Supplementary Material**.

Procedure

The participants individually attended three separate VR sessions; in each session, they were immersed in different VR software. The period between each session was 1 week for each participant (i.e., 3 weeks in total). The participants went through an induction pertinent to the VR software for that session and the specific HMD and controllers used (i.e., HTC Vive and its 6DoF wands-controllers) before being immersed. Subsequently, the participants were asked to play the respective VR game until they completed it, or they felt any discomfort or fatigue. The duration of each VR session was recorded from the time the software was started until the participant expressed that they wanted to discontinue. At the end of each session, participants were asked to complete the VRNQ. The "Job Simulator" was always used in the 1st session, "The Lab" was always used in the 2nd session, and "Rick and Morty: Virtual Rick-ality" was always used in the 3rd session.

Statistical Analyses

A reliability analysis of the VRNQ was conducted to calculate Cronbach's alpha and inspect whether the items have adequate internal consistency for research and clinical purposes. A Cronbach's alpha of 0.70–1.00 indicates good to excellent internal consistency (Nunally and Bernstein, 1994). A confirmatory factor analysis (CFA) was performed to examine the construct validity of the VRNQ in terms of convergent and discriminant validity (Cole, 1987). The reliability analysis and CFA were conducted using AMOS (version 24) (Arbuckle, 2014), and IBM Statistical Package for the Social Sciences (SPSS) 24.0 (Ibm Corp, 2016). Several tests for goodness of fit were implemented to allow the evaluation of VRNQ's structure. The (CFI), Tuckere Lewis index (TLI), standardized root mean square residual (SRMR), and the root mean squared error of approximation (RMSEA) were used to assess model fit. A CFI and TLI equal to or greater than 0.90 indicate good structural model fit to the data (Hu and Bentler, 1999; Jackson et al., 2009; Hopwood and Donnellan, 2010). An SRMR and RMSEA less than 0.08 postulate a good fit to the data (Hu and Bentler, 1999; Hopwood and Donnellan, 2010). Lastly, the variance of the results was assessed by dividing the χ^2 by the degrees of freedom (df), which is an indicator of the sample distribution (Hu and Bentler, 1999; Jackson et al., 2009; Hopwood and Donnellan, 2010).

The reliability and confirmatory factor analyses were conducted based on 120 observations (40 participants * 3 sessions with different software). The *a priori* sample size calculator for structural equation models was used to calculate the minimum sample size for model structure. This calculator uses the error function formula, the lower bound sample size formula for a structural equation model, and the normal distribution cumulative distribution function (Soper, 2019a), which are in perfect agreement with the recommendations for statistical power analysis for the behavioral sciences (Cohen, 2013). A sample size of 100 observations was suggested as the minimum for conducting CFA to examine the model structure with statistical power equal to or greater than 0.80. Hence, the 120 observations in our sample appear adequate to conduct a CFA with statistical power equal to or greater than 0.80.

Bayesian Pearson correlation analyses were conducted to examine whether any of the demographic variables were significantly associated with the VRNQ total score and sub-scores, or the length of the VR sessions. Bayesian paired samples *t*-tests were performed to investigate possible differences between each session's duration, as well as the VRNQ results for each VR game. Also, a Bayesian independent samples *t*-test examined whether there were any differences between gamers and non-gamers in the duration of the session. Lastly, a Bayesian linear regression was performed to examine the predictors of VRISE, where the Jeffreys–Zellner–Siow (JZS) mixed g-prior was used for the selection of the best model. JZS has the computational advantages of a g-prior in conjunction with the theoretical advantages of a Cauchy prior, which are valuable in variable selection for the best model (Liang et al., 2008; Rouder and Morey, 2012). For all the analyses, a Bayes Factor (BF_{10}) $\geq$ 10 was set for statistical inference, which indicates strong evidence in favor of the alternative hypothesis (Rouder and Morey, 2012; Wetzels and Wagenmakers, 2012; Marsman and Wagenmakers, 2017). All the Bayesian analyses were performed using JASP (Version 0.8.1.2) (Jasp Team, 2017). The Bayesian Pearson correlation analyses and Bayesian linear regression analysis were conducted based on 120 observations (40 participants * 3 different software sessions). The *post hoc* statistical power calculator was used to calculate the observed power of the best model using Bayesian linear regression analysis (Soper, 2019b).

RESULTS

Reliability Analysis and CFA

The reliability analysis demonstrated good to excellent Cronbach's α for each domain of the VRNQ (i.e., user experience – α = 0.89, game mechanics – α = 0.89, in-game assistance – α = 0.90, VRISE – α = 0.89; see **Table 2**), which indicate very good internal reliability (Nunally and Bernstein, 1994). VRNQ's fit indices are displayed in **Table 2** with their respective thresholds. The χ^2/df was 1.61, which indicates good

TABLE 2 | Internal reliability and goodness of fit for the VRNQ.

Statistics	Thresholds	Results
Cronbach's α	≥0.70	USER – 0.886 GM – 0.888 GA – 0.895 VR – 0.891
χ^2/df	≤2.00	1.610
Comparative fit index (CFI)	≥0.90	0.954
Tuckere Lewis index (TLI)	≥0.90	0.938
Standardized root mean square residual (SRMR)	<0.08	0.076
Root mean square error of approximation (RMSEA)	≤0.08	0.071

VRNQ Domains: USER, user experience; GM, game mechanics; GA, in-game assistance; VR VRISE.

variance in the sample (Hu and Bentler, 1999; Jackson et al., 2009; Hopwood and Donnellan, 2010). Both CFI and TLI were close to 0.95, which suggest a good fit for the VRNQ model (Hu and Bentler, 1999; Jackson et al., 2009; Hopwood and Donnellan, 2010). Comparably, SPMR and RMSEA values were between 0.06 and 0.08, which also support a good fit (Hu and Bentler, 1999; Jackson et al., 2009; Hopwood and Donnellan, 2010). The VRNQ's path diagram is displayed in **Figure 1**, where from left to right are depicted the correlations among the factors/domains of the VRNQ, the correlations between each factor/domain and its items, and the error terms for each item. The VRNQ items/questions are efficiently associated with their respective factor/domain, which shows good convergent validity (Cole, 1987). Furthermore, there was not any significant correlation amongst the factors/domains, which indicates good discriminant validity (Cole, 1987).

Descriptive Statistics of Sessions' Duration and VRNQ Scores

The descriptive statistics for the sessions' durations and the VRNQ scores are displayed in **Table 3**. In session 1, the participants were immersed for 59.65 (8.42) minutes. In session 1, the average time of gamers seems more than the average time of non-gamers (**Table 3**). In session 2, the participants spent 64.72 (6.24) minutes (**Table 3**). In session 3, gamers spent 70.44 (7.78) minutes, while non-gamers spent 65.73 (6.75) minutes (**Table 3**). The average total score of the VRNQ for all software was 126.30 (7.55) (maximum score is 140), where gamers and non-gamers scores did not appear to differ. Similarly, the median scores for each domain were 30–32 out of 35, where again gamers and non-gamers scores did not appear to differ. Importantly, all the VRISE scores (per item) for both gamers and non-gamers were equal to 5 (i.e., mild feeling), or 6 (i.e., very mild feeling), or 7 (absent feeling). The vast majority of scores were equal to 6 (i.e., very mild feeling) or 7 (absent feeling) (see **Figure 2**).

Minimum and Parsimonious Cut-Off Scores of VRNQ

Cut-off scores were calculated for the VRNQ total score and sub-scores to inspect the suitability of the assessed VR software (see **Table 4**). In the VRNQ, the ordinal 1–3 responses are paired with negative qualities, response 4 is paired with neutral/moderate qualities, and 5–7 responses are paired with positive qualities (see **Supplementary Material**). The minimum cut-offs suggest that if the median of the responses is 25 for every sub-score, and 100 in the total score (i.e., at least a median of 5 for every item), then the VRNQ outcomes indicate that the evaluated VR software is of an adequate quality not to cause any significant VRISE. Furthermore, the parsimonious cut-offs suggest that, if the median of the responses is 30 for every sub-score, and 120 for the total score (i.e., at least a median of 6 for every item) then the utilization of the parsimonious cut-offs more robustly supports the suitability of the VR software. The minimum and parsimonious cut-offs hence appear adequate to guarantee the safety, pleasantness, and appropriateness of the VR software for research and/or clinical purposes.

Bayesian *T*-Tests

The Bayesian independent samples *t*-test between gamers and non-gamers indicated that the former spent significantly more time in VR across the total duration for the 3 sessions (BF_{10} = 14.99), as well as the duration of the 1st session (BF_{10} = 2,532; see **Table 4**) (Wetzels and Wagenmakers, 2012; Marsman and Wagenmakers, 2017). The difference is much smaller in the total duration than the difference in the 1st session. Thus, the difference between the gamers and non-gamers in the total duration appears to be driven by the substantial difference in the1st session's duration (see **Table 5**). Conversely, the Bayesian paired samples *t*-test (i.e., differences between the VR games) indicated significant differences in the total score and every sub-score of VRNQ (see **Table 6**) between the VR software. The VR software in the 3rd session was evaluated higher than the VR software in the 1st and 2nd sessions, while the VR software in the 2nd session was rated better than the VR software in the 1st session. There was also an important difference between the duration of the 3rd session (longer) and the duration of the 1st session (shorter; BF_{10} = 103,568), while there was not a substantial difference between the duration of the 2nd and 3rd sessions (BF_{10} = 2.78), as well as between the duration of 1st and 2nd sessions (BF_{10} = 7.05; see **Table 6**) (Wetzels and Wagenmakers, 2012; Marsman and Wagenmakers, 2017).

Bayesian Pearson Correlation Analyses and Regression Analysis

The Bayesian Pearson correlation analyses did not show any significant correlation between age and any of the VRNQ scores, between age and duration of the sessions, between education and any of the VRNQ scores, or between education and duration of the sessions. However, the duration of the session was positively correlated with the total VRNQ score [BF_{10} = 81.54; $r(120) = 0.310$, $p < 0.001$]. Furthermore, the VRISE score substantially correlated with the following VRNQ items: immersion, pleasantness, graphics, sound, pick and place, tutorial's difficulty, tutorial's usefulness, tutorial's duration, instructions, and prompts (see **Table 7**). In contrast, VRISE did not significantly correlate with the following VRNQ

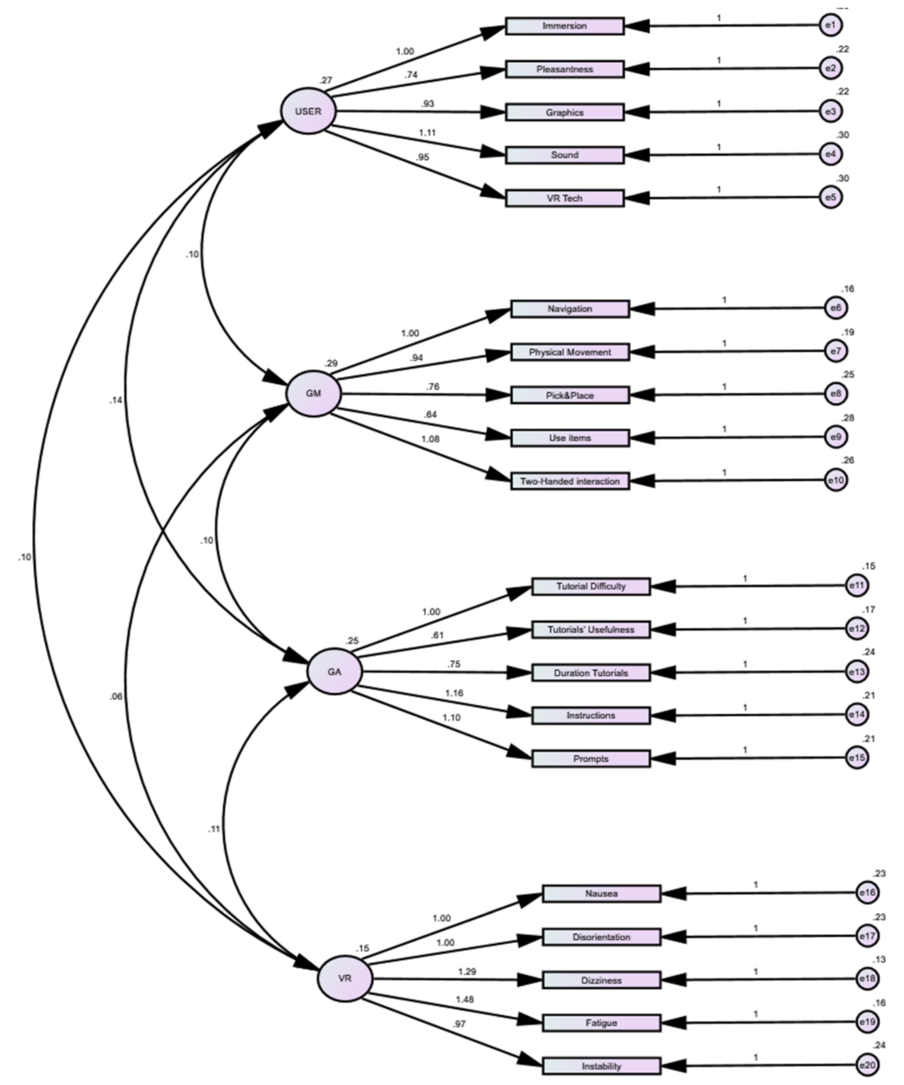


FIGURE 1 | CFA: model's path diagram. From left to right: the structural model illustrates the associations between VRNQ domains (paths with double headed arrow) and between each VRNQ domain and its items. At the right there are the error items (e) for each item; USER, user experience; GM, game mechanics; GA, in-game assistance; VR, VRISE.

items: VR tech, navigation, physical movement, use items, or two-handed interactions (see **Table 7**). Moreover, the Bayesian regression analysis indicated the five best models that predicted the VRNQ's VRISE score (see **Table 8**). The best model includes the following items from the VRNQ: immersion, graphics, sound, instructions, and prompts. All the predictors exceeded the prior inclusion probabilities (see **Figure 3**). The best model showed a $BF_M = 117.42$, whereas the second-best model displayed a $BF_M = 56.40$ (see **Table 8**); hence, the difference between the best model compared to the second-best model was robust (Rouder and Morey, 2012; Wetzels and Wagenmakers, 2012; Marsman and Wagenmakers, 2017). Also, the best model has an $R^2 = 0.324$ (see **Table 8**), which postulates that the model explains the 32.4% of the variance of VRISE score (Rouder and Morey, 2012; Wetzels and Wagenmakers, 2012). Lastly, the *post hoc* statistical power analysis for the best model indicated an observed statistical power of 0.998, $p < 0.001$, which postulates a high efficiency, precision, reproducibility, and reliability of the regression analysis and results (Button et al., 2013; Cohen, 2013).

TABLE 3 | Descriptive statistics: duration of VR sessions and VRNQ scores.

	Group	N	Mean (SD)	SE
Total duration	Gamers	18	199.39 (13.63)	3.21
	Non-Gamers	22	186.36 (11.76)	2.51
	Total	40	192.2 (14.09)	2.23
Duration of session 1	Gamers	18	65.61 (7.14)	1.68
	Non-Gamers	22	54.77 (5.91)	1.26
	Total	40	59.65 (8.42)	1.33
Duration of session 2	Gamers	18	63.33 (6.16)	1.45
	Non-Gamers	22	65.86 (6.21)	1.32
	Total	40	64.72 (6.24)	0.99
Duration of session 3	Gamers	18	70.44 (7.78)	1.83
	Non-Gamers	22	65.73 (6.75)	1.44
	Total	40	67.85 (7.52)	0.69
VRNQ total score out of 140 (across 3 sessions)	Gamers	18	127.2 (7.32)	0.99
	Non-Gamers	22	125.6 (7.71)	0.95
	Total	40	126.3 (7.55)	0.69
User's experience (across 3 sessions) out of 35	Gamers	18	31.37 (2.73)	0.34
	Non-Gamers	22	30.91 (2.73)	0.37
	Total	40	31.12 (2.73)	0.25
Game mechanics (across 3 sessions) out of 35	Gamers	18	31.50 (2.68)	0.37
	Non-Gamers	22	31.32 (2.61)	0.32
	Total	40	31.40 (2.63)	0.24
In-game assistance (across 3 sessions) out of 35	Gamers	18	31.70 (2.59)	0.35
	Non-Gamers	22	31.65 (2.52)	0.31
	Total	40	31.68 (2.54)	0.23
VRISE (across 3 sessions) out of 35	Gamers	18	32.67 (2.17)	0.30
	Non-Gamers	22	31.71 (2.56)	0.32
	Total	40	32.14 (2.43)	0.22

DISCUSSION

The VRNQ as a Research and Clinical Tool

The VRNQ is a short questionnaire (5–10 min administration time) which assesses the quality of VR software in terms of user experience, game mechanics, in-game assistance, and VRISE. The values of the fit indices of CFA (i.e., CFI, TLI, SPMR, and RMSEA) indicated that the VRNQ's structure was a good fit to the data, which postulates good construct validity for the VRNQ (Hu and Bentler, 1999; Jackson et al., 2009; Hopwood and Donnellan, 2010). In addition, the construct validity of the VRNQ was supported by its convergent and discriminant validity (Cole, 1987). VRNQ items were strongly correlated with their grouping factor, which indicates robust convergent validity, while there were substantially poor correlations between the factors, which postulates very good discriminant validity (Cole, 1987). Furthermore, the Cronbach's α for each VRNQ domain (i.e., user experience – $\alpha = 0.89$, game mechanics – $\alpha = 0.89$, in-game assistance – $\alpha = 0.90$, VRISE – $\alpha = 0.89$; see **Table 2**) suggest very good construct validity (Nunally and Bernstein, 1994). Henceforth, the VRNQ emerges as a valid and suitable tool to evaluate the quality of the VR research/clinical software as well as the intensity of the adverse VRISE.

Furthermore, minimum and parsimonious cut-off scores were calculated for the VRNQ total score and sub-scores to inspect the suitability of the assessed VR software. The minimum cut-offs indicate the lowest acceptable quality that VR research/clinical software should be, while the parsimonious cut-offs are offered for more robust support of the VR software's suitability, which may be required in experimental and clinical designs with more conservative standards. However, the individual scores from the VRNQ may be modulated by individual differences and preferences unrelated to the quality of the software (Kortum and Peres, 2014). In addition, the VRNQ produces ordinal data; therefore, the median is the appropriate measure for their analysis (Harpe, 2015). Hence, the median VRNQ scores for the whole sample should be used to assess the VR software's quality effectively. Also, the medians of the VRNQ total score and sub-scores allow the generalization of the results and comparison between different VR software (Kortum and Peres, 2014; Harpe, 2015). Researchers, clinicians, and/or research software developers should use the medians of the VRNQ total score and sub-scores to assess whether the implemented VR software exceed the minimum or parsimonious cut-offs. Hence, if the medians of the VRNQ sub-scores and totals score for VR research software meet the minimum cut-offs, then these results support the VR software's suitability. Likewise, if the medians of VRNQ sub-scores and totals score for VR research software meet the parsimonious cut-offs, then these results provide even stronger support for its suitability. However, median scores below these

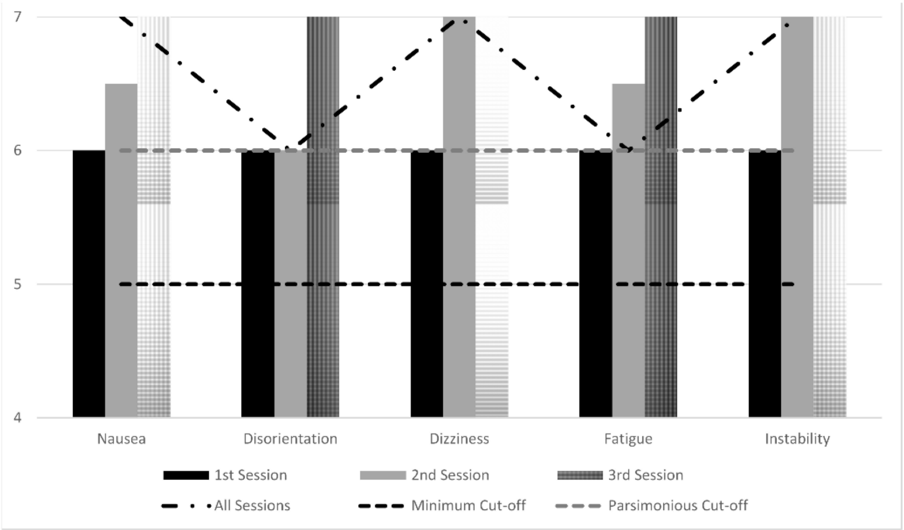


FIGURE 2 | VRISE intensity in VR sessions as measured by VRNQ. Median scores of VRISE items of VRNQ; VRNQ Minimum Cut-off (≥); VRNQ Parsimonious Cut-off (≥); 1, Extreme intense feeling; 2, Very intense feeling; 3, Intense feeling; 4, Moderate feeling; 5, Mild feeling; 6, Very mild feeling; 7, Absent feeling.

cut-offs suggest that the suitability of the VR software is questionable, but they do not indicate that this VR software is certainly unsuitable.

Also, VRNQ appears as an appropriate tool to measure both VRISE and VR software features compared to other questionnaires. The SSQ is the most implemented questionnaire in VR studies. However, the SSQ only considers the symptoms pertinent to simulator sickness and it does not assess software attributes (Kennedy et al., 1993), while there is a dispute that simulator sickness symptomatology may not be the same as VRISE (Stanney et al., 1997). Alternatively, Virtual reality sickness questionnaire (VRSQ) was recently developed (Kim et al., 2018). The development of VRSQ was based on the SSQ, where the researchers attempted to isolate the items which are pertinent to VRISE (Kim et al., 2018). However, their sample size was relatively small (i.e., 24 participants * 4 sessions = 96 observations) (Kim et al., 2018). Notably, the factor analyses of Kim et al. (2018) accepted only items pertinent to oculomotor and disorientation components of SSQ, and rejected all the items pertinent to nausea (i.e., 7 items) (Kim et al., 2018), while nausea is the most frequent symptom in VRISE (Stanney et al., 1997; Sharples et al., 2008; Bohil et al., 2011; de França and Soares, 2017; Palmisano et al., 2017). Also, comparable to SSQ, VRSQ does not consider software features. Hence, the VRNQ appears to be the only valid and suitable tool to evaluate both the intensity of predominant VRISE and the quality of VR software features.

The VRNQ allows researchers to report the quality of VR software and/or the intensity of VRISE in their VR studies. However, an in-depth assessment of the numerous software features requires a questionnaire with more than the 20 questions of the VRNQ (Zarour et al., 2015). For an in-depth software analysis, questionnaires with more questions pertinent to the whole spectrum of software features should be preferred (Zarour et al., 2015). Additionally, the VRNQ has solely five items pertinent to VRISE. Hence, it does not offer an exhaustive assessment of VRISE. Studies that aim to investigate VRISE in depth should opt for a tool which contains more items pertinent to VRISE than VRNQ (e.g., SSQ). The VNRQ is a brief questionnaire (5–10 min administration time) including 20 items, which enables researchers, clinicians, and research software developers to evaluate and report the quality of the VR software and the intensity of VRISE for research and clinical purposes.

Maximum Duration of VR Sessions

The duration of the VR session is a crucial factor in research and/or clinical design. In our sample, the participants discontinued the VR session due to loss of interest, while none discontinued due to VRISE. In the 1st session, gamers spent significantly more time immersed than the non-gamers; a difference which modulated the difference between the two

TABLE 4 | VRNQ cut-offs.

Score	Minimum cut-offs	Parsimonious cut-offs
User experience	≥25/35	≥30/35
Game mechanics	≥25/35	≥30/35
In-game assistance	≥25/35	≥30/35
VRISE	≥25/35	≥30/35
VRNQ total score	≥100/140	≥120/140

The median of each sub-score and totals scores should meet the suggested cut-offs to support that the evaluated VR software has an adequate quality without any significant VRISE. The utilization of the parsimonious cut-offs more robustly supports the suitability of the VR software.

TABLE 5 | Bayesian independent samples *t*-test: gamers against non-gamers.

Variables	Significance	BF10	Error %
Age		0.323	0.006
Education		0.325	0.006
Total duration	*	14.987	7.044e-6
Session 1 duration	***	2531.886	7.491e-8
Session 2 duration		0.595	0.006
Session 3 duration		1.580	0.003
VRNQ total		0.425	0.007
User's experience		0.359	0.006
Game mechanics		0.315	0.006
In-game assistance		0.315	0.006
VRISE		0.745	0.003

*BF_{10} = Bayes Factor; * $BF_{10} > 10$, *** $BF_{10} > 100$.*

groups in the summed duration across all sessions. However, it is worth noting that there was not a significant difference between the two groups in the time spent in VR for the 2nd and 3rd sessions. The observed difference in the 1st session and the absence of a difference in the later sessions' durations postulates that when users are familiarized with the VR technology, while the influence of their gaming experience on the session's duration becomes insignificant. In support of this, a recent study showed that user gaming experience does not affect the perceived workload of the users in VR (Lum et al., 2018). Hence, the level of familiarization of the participants with the VR technology appears to affect substantially the duration of the VR session.

Nevertheless, in the whole sample, irrespective of participants' gaming experience, the durations of the 2nd and 3rd sessions are sufficiently longer than the duration of the 1st session. The duration of the 3rd session is not significantly longer than the duration of the 2nd session. Furthermore, given that in each session, a different VR software was administered, the VRNQ correspondingly pinpointed significant differences amongst the implemented VR software' quality. All the VRNQ scores for the 3rd session's VR software are greater than the 2nd session's VR software scores. Similarly, all the VRNQ scores for the 2nd session's VR software are greater than the 1st session's VR software scores. Also, the duration of VR session was positively correlated with the total score of VRNQ. Thus, the quality of the VR software as measured by the VRNQ seems to be significantly associated with the duration of the VR session.

Overall, in every session, the intensity of VRISE was reported as very mild to absent by the vast majority of the sample. However, comparable to the rest of the VRNQ scores, the VRISE score for the 3rd VR session was significantly higher (i.e., milder feeling) than the 2nd and 3rd sessions. Similarly, the VRISE score for the 2nd session's VR software was substantially higher than the 1st session's VR software score. Notably, there was not any difference between gamers and non-gamers in the VRNQ scores across the three sessions. Equally, the age and education of participants did not correlate with any of the VRNQ scores or the duration of sessions. Thus, the age, education, and gaming experience of the participants did not affect the responses in the VRNQ. Therefore, the observed differences in the VRISE scores between the VR sessions support that the quality of the VR software as measured by the VRNQ and the level of familiarization of the participants with the VR technology also affect the intensity of VRISE.

The findings postulate that the implementation of VR software with a maximum duration between 55 and 70 min is substantially feasible. However, long exposures in VR have been found to increase the probability of experiencing VRISE and the intensity of VRISE (Sharples et al., 2008). In our sample, especially in the 3rd session, which was substantially longer than the other sessions, the intensity of VRISE was significantly lower than the rest of the sessions. As discussed above, the substantially lower intensity of VRISE in the 3rd session appears to be a result of increased VR familiarity, and the better quality of the implemented VR software as measured by the VRNQ. Hence, researchers and/or clinicians should consider the quality of their VR software to define the appropriate duration of their VR session. In research and clinical designs where the duration of the VR session is required to be between 55 and 70 min, the researchers and/or clinicians should opt for the parsimonious cut-offs of the VRNQ to ensure adequate quality of their VR software to facilitate longer sessions without significant VRISE. Additionally, an extended introductory tutorial which allows participants to familiarize themselves with the VR technology and mechanics would assist with the implementation of longer (i.e., 55–70 min) VR sessions, where the presence and intensity of VRISE would not be significant.

The Quality of VR Software and VRISE

The VRISE score substantially correlated with almost every item under the section of user experience and in-game assistance (see **Table 6**). However, the VRISE score did not correlate with VR tech (the item under the user experience's domain) or most of the items under the section of game mechanics. The quality of VR hardware (i.e., the HMD and its controllers) and interactions (i.e., ergonomic or non-ergonomic) with the virtual environment are crucial for the alleviation or evasion of VRISE (Kourtesis et al., 2019). Nevertheless, in this sample, the VR tech item (i.e., the quality of the internal and external VR hardware) was not expected to correlate with the VRISE score, because the HMD and its 6DoF controllers were the

TABLE 6 | Bayesian paired samples *t*-tests: differences between the VR software.

Pairs		Significance	BF10	Error %
Session 2 duration	Session 1 duration		7.049	∼0.001
Session 3 duration	Session 2 duration		2.783	∼3.276e-4
Session 3 duration	Session 1 duration	***	103568.858	NaN
S3 VRNQ total	S2 VRNQ total	***	6.942e + 12	NaN
S3 VRNQ total	S1 VRNQ total	***	3.520e + 20	NaN
S2 VRNQ total	S1 VRNQ total	***	8.500e + 17	NaN
S3 VRISE	S2 VRISE	***	22075.036	NaN
S3 VRISE	S1 VRISE	***	1.322e + 10	NaN
S2 VRISE	S1 VRISE	***	1.160e + 7	NaN
S3 in-game assistance	S2 in-game assistance	***	207216.904	NaN
S2 in-game assistance	S1 in-game assistance	***	1.197e + 7	NaN
S3 in-game assistance	S1 in-game assistance	***	8.028e + 10	NaN
S3 game mechanics	S2 game mechanics	***	274310.417	NaN
S2 game mechanics	S1 game mechanics	***	4.883e + 14	NaN
S3 game mechanics	S1 game mechanics	***	2.876e + 14	NaN
S3 user's experience	S2 user's experience	***	2.873e + 7	NaN
S3 in-game assistance	S1 user's experience	***	2.597e + 7	NaN
S2 user's experience	S1 user's experience	***	1.708e + 6	NaN

*BF_{10} = Bayes Factor; * $BF_{10} > 10$, ** $BF_{10} > 30$, *** $BF_{10} > 100$; S1, Session 1; S2, Session 2; S3, Session 3.*

same for all 3 VR software versions and sessions. Hence, the variance in the responses to this item was limited. Also, the three VR software games share common game mechanics, especially the same navigation system (i.e., teleportation) and a similar amount of physical mobility. Likewise, apart from some controls (i.e., the button to grab items), the interaction systems of the implemented VR software were very proximal. Therefore, the absence of a correlation between VRISE scores and most of the items in the game mechanics' section was also an expected outcome. Nonetheless, the VRISE score was strongly associated with the level of immersion and enjoyment, the quality of graphics and sound, the comfort to pick and place 3D objects, and the usefulness of in-game assistance modes (i.e., tutorials, instructions, and prompts).

TABLE 7 | Bayesian Pearson correlations analyses: VRISE score with VRNQ items.

Pairs		Significance	BF10	r
VRISE	Immersion	***	1226.538	0.371
VRISE	Pleasantness	*	20.504	0.273
VRISE	Graphics	***	1629.195	0.377
VRISE	Sound	***	18586.578	0.421
VRISE	VR Tech		5.094	0.228
VRISE	Navigation		4.808	0.226
VRISE	Physical movement		2.229	0.197
VRISE	Pick and place	***	175.087	0.329
VRISE	Use items		0.405	0.109
VRISE	Two-handed interaction		0.506	0.123
VRISE	Tutorial difficulty	***	28252.587	0.428
VRISE	Tutorials usefulness	***	161.949	0.327
VRISE	Tutorials' duration	***	128.539	0.322
VRISE	Instructions	***	952.871	0.366
VRISE	Prompts	***	706510.726	0.476

*BF_{10} = Bayes Factor; * $BF_{10} > 10$, ** $BF_{10} > 30$, *** $BF_{10} > 100$;*

The items which correlated with the VRISE score were also included in the best models of predicting its value (see **Table 7**). Importantly, the best model includes as predictors of VRISE, the level of immersion, the quality of graphics and sound, and the helpfulness of in-game instructions and prompts (see **Table 7**). The higher scores for prompts and instructions indicate that the user was substantially assisted by the in-game assistance (e.g., an arrow showing the direction that the user should follow) to orientate and guide his or herself from one point of interest to the next in accordance with the scenario of the VR experience. This may be interpreted as ease to orient and interact with the virtual environment, as well as a significant decrease in confusion (Brade et al., 2018). The quality of the in-game assistance methods is essential for the usability and enjoyment that VR software offers (Brade et al., 2018). Equally, the quality of the graphics is predominantly dependent upon rendering which encompasses the in-game quality of the image known as perceptual quality, and the exclusion of redundant visual information known as occlusion culling (Lavoué and Mantiuk, 2015). The improvement of these two factors not only results in improved quality of the graphics but also in improved performance of the software (Brennesholtz, 2018). Furthermore, the spatialized sound of VR software, which assists the user to orient his or herself (Ferrand et al., 2017), deepens the experienced immersion (Riecke et al., 2011), and enriches the geometry of the virtual space without affecting the performance of the software (Kobayashi et al., 2015). Lastly, the level of immersion appears to be negatively correlated with the frequency and intensity of VRISE (Milleville-Pennel and Charron, 2015; Weech et al., 2019). The best model hence aligns with the relevant

TABLE 8 | Models' comparison: predictors of VRISE score.

Models	P(M)	P(M| data)	BF_M	BF_{10}	R^2
Prompts + Sound + Graphics + Immersion + Instructions	0.004	0.304	117.42***	1.000	0.324
Prompts + Graphics + Immersion + Instructions + Pleasantness	0.004	0.173	56.47**	0.571	0.317
Prompts + Sound + Graphics + Immersion + Instructions + Pick and Place	0.004	0.161	43.15*	0.443	0.330
Prompts + Sound + Graphics + Immersion + Instructions + Pick and Place Tutorials Usefulness + Pleasantness	0.021	0.123	6.62	0.072	0.337
Prompts + Graphics + Immersion + Instructions + Pick and Place + Tutorials Usefulness + Pleasantness	0.008	0.077	10.72*	0.121	0.329

*P, Probability; M, Model; BF_M, Model's Bayesian Factor; * $BF_M > 10$, ** $BF_M > 30$, *** $BF_M > 100$; BF_{10} = BF against null model.*

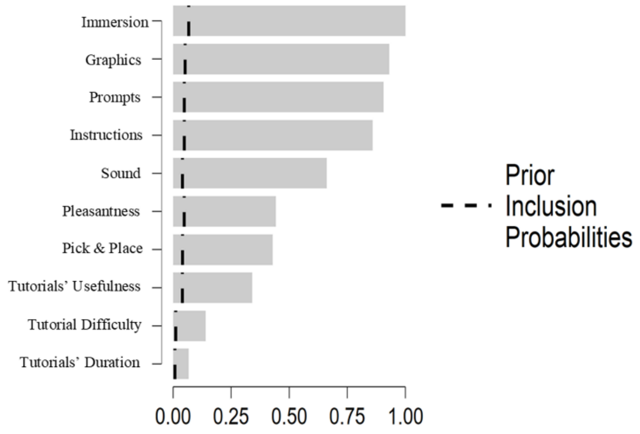


FIGURE 3 | Variables' prior inclusion Probabilities.

literature and provides further evidence in support of the utility of the VRNQ as a valid and efficient tool to appraise the quality of the VR software and intensity of VRISE.

Limitations and Future Studies

This study also has some limitations. In this study, construct validity for the VRNQ is provided. However, future work should endeavor to provide convergent validation of the VRNQ with tools that measure VRISE symptomatology (e.g., SSQ) and/or VR software attributes. Moreover, the sample size was relatively small, but it offered an adequate statistical power for the conducted analyses. Also, the VRNQ does not directly quantify linear or angular accelerations, which may induce intense VRISE in a relatively short period of time (McCauley and Sharkey, 1992; LaViola, 2000; Gavgani et al., 2018). However, the VRNQ quantifies the effect(s) of linear and angular accelerations (i.e., VRISE), where VR software with a highly provocative content (e.g., linear and angular accelerations) would fail to meet or exceed the VRNQ cut-offs for the VRISE domain. Furthermore, the study utilized only one type of VR hardware, which did not allow us to inspect the effect of VR HMD's quality on VRISE presence and intensity. Similarly, our VR software did not allow us to compare different ergonomic interactions or levels of provocative potency pertaining to VRISE. Future studies with a larger sample, various types of VR hardware, and VR software with substantially more diverse features will offer further insights on the impact of software features on VRISE intensity, as well as provide additional support for the VRNQ's structural model. Lastly, neuroimaging (e.g., electroencephalography) and physiological data (e.g., heart rates) may correlate, classify, and predict VRISE symptomatology (Kim et al., 2005; Dennison et al., 2016, 2019). Hence, future studies should consider collecting neuroimaging and/or physiological data that could further elucidate the relationship between VRNQ's VRISE score(s) and brain region activation or cardiovascular responses (e.g., heart rate).

CONCLUSION

This study showed that the VRNQ is a valid and reliable tool which assesses the quality of VR software and intensity of VRISE. Our findings support the viability of VR sessions with a duration up to 70 min, when the participants are familiarized with VR tech through an induction session, and the quality of the VR software meets the parsimonious cut-offs of VRNQ. Also, our results offered insights on the software-related predictors of VRISE intensity, such as the level of immersion, the quality of graphics and sound, and the helpfulness of in-game instructions and prompts. Finally, the VRNQ enables researchers to quantitatively assess and report the quality of VR software features and intensity of VRISE, which are vital for the efficacious implementation of immersive VR systems in cognitive neuroscience and neuropsychology. The minimum and parsimonious cut-offs of VRNQ may appraise the suitability of VR software for implementation in research and clinical settings. The VRNQ and the findings of this study contribute to the endeavor of establishing thorough VR research and clinical methods that are crucial to guarantee the viability of implementing immersive VR systems in cognitive neuroscience and neuropsychology.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Philosophy, Psychology and Language Sciences Research Ethics Committee of the University of Edinburgh. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PK had the initial idea and contributed to every aspect of this study. SC, LD, and SM contributed to the methodological aspects and the discussion of the results. The VRNQ may be downloaded from **Supplementary Material.**

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00417/full#supplementary-material

REFERENCES

Arafat, I. M., Ferdous, S. M. S., and Quarles, J. (2018). "Cybersickness-provoking virtual reality alters brain signals of persons with multiple sclerosis," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, (Reutlingen: IEEE), 1–120.

Arbuckle, J. L. (2014). *Amos (Version 23.0)*. Chicago: IBM SPSS.

Bohil, C. J., Alicea, B., and Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* 12, 752–762. doi: 10.1038/nrn3122

Brade, J., Dudczig, M., and Klimant, P. (2018). "Using virtual prototyping technologies to evaluate human-machine-interaction concepts," in *aw&I Conference*, Vol. 3, (Chemnitz).

Brennesholtz, M. S. (2018). 3—1: invited paper: vr standards and guidelines. *SID Symposium Dig. Tech. Pap.* 49, 1–4. doi: 10.1002/sdtp.12476

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., et al. (2013). Confidence and precision increase with high statistical power. *Nat. Rev. Neurosci.* 14:585. doi: 10.1038/nrn3475-c4

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences.*New York, NY: Routledge.

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *J. Consul. Clin. Psychol.* 55, 584–594. doi: 10.1037//0022-006x.55.4.584

de França, A. C. P., and Soares, M. M. (2017). "Review of virtual reality technology: an ergonomic approach and current challenges," in *International Conference on Applied Human Factors and Ergonomics*, (Cham: Springer), 52–61. doi: 10.1007/978-3-319-60582-1_6

Dennison, M. Jr., D'Zmura, M., Harrison, A., Lee, M., and Raglin, A. (2019). "Improving motion sickness severity classification through multi-modal data fusion," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, Vol. 11006, (Baltimore, MA.), 110060T.

Dennison, M. S., Wisti, A. Z., and D'Zmura, M. (2016). Use of physiological signals to predict cybersickness. *Displays* 44, 42–52. doi: 10.1016/j.displa.2016.07.002

Ferrand, S., Alouges, F., and Aussal, M. (2017). "Binaural spatialization methods for indoor navigation," in *Audio Engineering Society Convention*, Vol. 142, (New York, NY: Audio Engineering Society).

Gavgani, A. M., Wong, R. H., Howe, P. R., Hodgson, D. M., Walker, F. R., and Nalivaiko, E. (2018). Cybersickness-related changes in brain hemodynamics: a pilot study comparing transcranial Doppler and near-infrared spectroscopy assessments during a virtual ride on a roller coaster. *Physiol. Behav.* 191, 56–64. doi: 10.1016/j.physbeh.2018.04.007

Harpe, S. E. (2015). How to analyze likert and other rating scale data. *Curr. Pharm. Teach. Learn.* 7, 836–850. doi: 10.1016/j.cptl.2015.08.001

Hopwood, C. J., and Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personal. Soc. Psychol. Rev.* 14, 332–346. doi: 10.1177/1088868310361240

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struc. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118

Ibm Corp. (2016). *IBM SPSS Statistics for Windows, Version 24.0.* Armonk, NY: IBM Corp.

Jackson, D. L., Gillaspy, J. A. Jr., and Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychol. Methods* 14, 6–23. doi: 10.1037/a0014694

Jasp Team, (2017). *JASP (Version 0.8.1.2).*

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* 3, 203–220. doi: 10.1207/s15327108ijap0303_3

Kim, H. K., Park, J., Choi, Y., and Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. *Appl. Ergon.* 69, 66–73. doi: 10.1016/j.apergo.2017.12.016

Kim, Y. Y., Kim, H. J., Kim, E. N., Ko, H. D., and Kim, H. T. (2005). Characteristic changes in the physiological components of cybersickness. *Psychophysiology* 42, 616–625.

Kobayashi, M., Ueno, K., and Ise, S. (2015). The effects of spatialized sounds on the sense of presence in auditory virtual environments: a psychological and physiological study. *Presence* 24, 163–174. doi: 10.1162/pres_a_00226

Kortum, P., and Peres, S. C. (2014). The relationship between system effectiveness and subjective usability scores using the system usability scale. *Int. J. Hum. Comput. Int.* 30, 575–584. doi: 10.1080/10447318.2014.904177

Kourtesis, P., Collina, S., Doumas, L. A. A., and MacPherson, S. E. (2019). Technological competence is a precondition for effective implementation of virtual reality head mounted displays in human neuroscience: a technological review and meta-analysis. *Front. Hum. Neurosci.* 13:342.

LaViola, J. J. Jr. (2000). A discussion of cybersickness in virtual environments. *ACM Sigchi Bull.* 32, 47–56. doi: 10.1145/333329.333344

Lavoué, G., and Mantiuk, R. (2015). "Quality assessment in computer graphics," in *Visual Signal Quality Assessment*, (Cham: Springer), 243–286. doi: 10.1007/978-3-319-10368-6_9

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *J. Am. Stat. Assoc.* 103, 410–423.

Lum, H. C., Greatbatch, R., Waldfogle, G., and Benedict, J. (2018). "How Immersion, Presence, Emotion, & Workload Differ in Virtual Reality and Traditional Game Mediums," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 62, (Los Angeles, CA: SAGE Publications), 1474–1478. doi: 10.1177/1541931218621334

Marsman, M., and Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *Eur. J. Dev. Psychol.* 14, 545–555. doi: 10.1080/17405629.2016.1259614

McCauley, M. E., and Sharkey, T. J. (1992). Cybersickness: perception of self-motion in virtual environments. *Presence* 1, 311–318. doi: 10.1162/pres.1992.1.3.311

Milleville-Pennel, I., and Charron, C. (2015). Do mental workload and presence experienced when driving a real car predispose drivers to simulator sickness? An exploratory study. *Accid. Anal. Prev.* 74, 192–202. doi: 10.1016/j.aap.2014.10.021

Mittelstaedt, J., Wacker, J., and Stelling, D. (2018). Effects of display type and motion control on cybersickness in a virtual bike simulator. *Displays* 51, 43–50. doi: 10.1016/j.displa.2018.01.002

Mittelstaedt, J. M., Wacker, J., and Stelling, D. (2019). VR aftereffect and the relation of cybersickness and cognitive performance. *Vir. Real.* 23, 143–154. doi: 10.1007/s10055-018-0370-3

Nalivaiko, E., Davis, S. L., Blackmore, K. L., Vakulin, A., and Nesbitt, K. V. (2015). Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiol. Behav.* 151, 583–590. doi: 10.1016/j.physbeh.2015.08.043 doi: 10.1016/j.physbeh.2015.08.043

Nesbitt, K., Davis, S., Blackmore, K., and Nalivaiko, E. (2017). Correlating reaction time and nausea measures with traditional measures of cybersickness. *Displays* 48, 1–8. doi: 10.1016/j.displa.2017.01.002

Nunally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory*, 3rd Edn. New Yokr, NY: Mcgraw-Hill.

Palmisano, S., Mursic, R., and Kim, J. (2017). Vection and cybersickness generated by head-and-display motion in the oculus rift. *Displays* 46, 1–8. doi: 10.1016/j.displa.2016.11.001

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front. Hum. Neurosci.* 9:660. doi: 10.3389/fnhum.2015.00660

Parsons, T. D., McMahan, T., and Kane, R. (2018). Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms. *Clin. Neuropsychol.* 32, 16–41. doi: 10.1080/13854046.2017.1337932

Riecke, B. E., Feuereissen, D., Rieser, J. J., and McNamara, T. P. (2011). "Spatialized sound enhances biomechanically-induced self-motion illusion (vection)," in *Proceedings of the SIGCsSHI Conference on Human Factors in Computing Systems*, (New York, NY: ACM.), 2799–2802.

Rouder, J. N., and Morey, R. D. (2012). Default bayes factors for model selection in regression. *Multivar. Behav. Res.* 47, 877–903. doi: 10.1080/00273171.2012.734737

Sharples, S., Cobb, S., Moody, A., and Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. *Displays* 29, 58–69. doi: 10.1016/j.displa.2007.09.005

Soper, D. S. (2019a). *A-priori Sample Size Calculator for Structural Equation Models*. Available at http://www.danielsoper.com/statcalc (accessed July 10, 2019).

Soper, D. S. (2019b). *Post-hoc Statistical Power Calculator for Multiple Regression*. Available at http://www.danielsoper.com/statcalc (accessed July 10, 2019).

Stanney, K. M., Kennedy, R. S., and Drexler, J. M. (1997). "Cybersickness is not simulator sickness," In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, (Los Angeles, CA: SAGE Publications). 41, 1138-1142.

Toschi, N., Kim, J., Sclocco, R., Duggento, A., Barbieri, R., Kuo, B., et al. (2017). Motion sickness increases functional connectivity between visual motion and nausea-associated brain regions. *Auton. Neurosci.* 202, 108–113. doi: 10.1016/j.autneu.2016.10.003

Weech, S., Kenny, S., and Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: a review. *Front. Psychol.* 10:158. doi: 10.3389/fpsyg.2019.00158

Wetzels, R., and Wagenmakers, E. J. (2012). A default bayesian hypothesis test for correlations and partial correlations. *Psychon. Bull. Rev.* 19, 1057–1064. doi: 10.3758/s13423-012-0295-x

Zarour, M., Abran, A., Desharnais, J. M., and Alarifi, A. (2015). An investigation into the best practices for the successful design and implementation of lightweight software process assessment methods: a systematic literature review. *J. Syst. Softw.* 101, 180–192. doi: 10.1016/j.jss.2014.11.041

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Chapter 3 provided the validation of VRNQ and the possible maximum duration of a VR session. The VRNQ hence may be now implemented to assess the software features of the VR-EAL, as well as the intensity of VRISE that the VR-EAL induces. Also, Chapter 3 identified the preconditions (e.g., familiarisation with the VR systems) that should be met in order the VR session to be long enough for an extensive neuropsychological assessment. This identification will assist with the design (e.g., an alternating scenario with both tutorials and tasks) and the development of VR-EAL (e.g., required software features).

Chapter 4: Guidelines for the Development of Immersive Virtual Reality Software for Cognitive Neuroscience and Neuropsychology: The Development of Virtual Reality Everyday Assessment Lab (VR-EAL), a Neuropsychological Test Battery in Immersive Virtual Reality

Chapter 2 indicated which VR HMDs should be used to avoid or mitigate VRISE, as well as which VR software characteristics assist with the mitigation of VRISE. Chapter 3 showed the possible maximum duration of the VR sessions and validated the VRNQ which assesses both the quality of VR software features and the intensity of VRISE. In this chapter, the development of VR-EAL will be performed based on findings of the previous Chapters. Also, different versions of VR-EAL will be developed and assessed using the VRNQ. The version of VR-EAL that will be accepted for implementation should meet or surpass the parsimonious cut-offs of the VRNQ to ensure a substantial mitigation of VRISE frequency and intensity.

However, the development of an immersive VR neuropsychological assessment of everyday cognitive functions requires competency in programming, software development, and psychometrics. These multidisciplinary requirements may discourage the adoption of immersive VR methods in cognitive neuroscience and neuropsychology. For this reason, Chapter 4 will provide guidelines for the development of an immersive VR neuropsychological assessment by presenting and discussing the software development process of the VR-EAL.

This chapter has been published in the Frontiers in Computer Science: Kourtesis, P., Korre, D., Collina, S., Doumas, L. A., & MacPherson, S. E. (2020). Guidelines for the development of immersive virtual reality software for cognitive neuroscience and neuropsychology: the development of virtual reality everyday assessment lab (VR-EAL), a neuropsychological test battery in immersive virtual reality. Frontiers in Computer Science, 1, 12.doi: https://doi.org/10.3389/fcomp.2019.00012

frontiers
in Computer Science

ORIGINAL RESEARCH
published: 14 January 2020
doi: 10.3389/fcomp.2019.00012

Guidelines for the Development of Immersive Virtual Reality Software for Cognitive Neuroscience and Neuropsychology: The Development of Virtual Reality Everyday Assessment Lab (VR-EAL), a Neuropsychological Test Battery in Immersive Virtual Reality

Panagiotis Kourtesis [1,2,3,4]*, **Danai Korre** [5], **Simona Collina** [3,4], **Leonidas A. A. Doumas** [2] and **Sarah E. MacPherson** [1,2]

[1] Human Cognitive Neuroscience, Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom, [2] Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom, [3] Lab of Experimental Psychology, Suor Orsola Benincasa University of Naples, Naples, Italy, [4] Interdepartmental Centre for Planning and Research "Scienza Nuova", Suor Orsola Benincasa University of Naples, Naples, Italy, [5] Centre for Intelligent Systems and Their Applications, School of Informatics, University of Edinburgh, Edinburgh, United Kingdom

Virtual reality (VR) head-mounted displays (HMD) appear to be effective research tools, which may address the problem of ecological validity in neuropsychological testing. However, their widespread implementation is hindered by VR induced symptoms and effects (VRISE) and the lack of skills in VR software development. This study offers guidelines for the development of VR software in cognitive neuroscience and neuropsychology, by describing and discussing the stages of the development of Virtual Reality Everyday Assessment Lab (VR-EAL), the first neuropsychological battery in immersive VR. Techniques for evaluating cognitive functions within a realistic storyline are discussed. The utility of various assets in Unity, software development kits, and other software are described so that cognitive scientists can overcome challenges pertinent to VRISE and the quality of the VR software. In addition, this pilot study attempts to evaluate VR-EAL in accordance with the necessary criteria for VR software for research purposes. The VR neuroscience questionnaire (VRNQ; Kourtesis et al., 2019b) was implemented to appraise the quality of the three versions of VR-EAL in terms of user experience, game mechanics, in-game assistance, and VRISE. Twenty-five participants aged between 20 and 45 years with 12–16 years of full-time education evaluated various versions of VR-EAL. The final version of VR-EAL achieved high scores in every sub-score of the VRNQ and exceeded its parsimonious cut-offs. It also appeared to have better in-game assistance and game mechanics, while its improved graphics substantially

Edited by:
Ioanna Iacovides,
University of York, United Kingdom

Reviewed by:
Manuela Chessa,
University of Genoa, Italy
Magdalena Mendez-Lopez,
University of Zaragoza, Spain
Lukas Gehrke,
Technische Universität
Berlin, Germany

***Correspondence:**
Panagiotis Kourtesis
pkourtes@exseed.ed.ac.uk

Specialty section:
This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Computer Science

Received: 11 September 2019
Accepted: 16 December 2019
Published: 14 January 2020

Citation:
Kourtesis P, Korre D, Collina S, Doumas LAA and MacPherson SE (2020) Guidelines for the Development of Immersive Virtual Reality Software for Cognitive Neuroscience and Neuropsychology: The Development of Virtual Reality Everyday Assessment Lab (VR-EAL), a Neuropsychological Test Battery in Immersive Virtual Reality. Front. Comput. Sci. 1:12. doi: 10.3389/fcomp.2019.00012

increased the quality of the user experience and almost eradicated VRISE. The results substantially support the feasibility of the development of effective VR research and clinical software without the presence of VRISE during a 60-min VR session.

Keywords: virtual reality, prospective memory, episodic memory, cybersickness, executive function, neuropsychology, everyday functioning, attention

INTRODUCTION

In cognitive neuroscience and neuropsychology, the collection of cognitive and behavioral data is predominantly achieved by implementing psychometric tools (i.e., cognitive screening and testing). The psychometric tools are principally limited to paper-and-pencil and computerized (i.e., 2D and 3D applications) forms. Psychometric tools in clinics and/or laboratories display several limitations and discrepancies between the observed performance in the laboratory/clinic and the actual performance of individuals in everyday life (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015). The functional and predictive association between an individual's performance on a set of neuropsychological tests and the individual's performance in various everyday life settings is called ecological validity. Ecological validity is considered an important issue that cannot be resolved by the currently available assessment tools (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015).

Ecological validity is especially important in the assessment of certain cognitive functions, which are crucial for performance in everyday life (Chaytor and Schmitter-Edgecombe, 2003). In particular, executive functioning (e.g., multitasking, planning ability, and mental flexibility) has been found to predict occupational and academic success (Burgess et al., 1998). Similarly, the ecologically valid measurement of memory (e.g., episodic memory) and attentional processes (e.g., selective, divided, and sustained attention) have been seen as predictors of overall performance in everyday life (Higginson et al., 2000). Lastly, prospective memory (i.e., the ability to remember to carry out intended actions at the correct point in the future; McDaniel and Einstein, 2007) plays an important role in everyday life and the assessment of prospective memory abilities requires ecologically valid tasks (Phillips et al., 2008).

Current ecologically valid tests are not thought to encompass the complexity of real-life situations (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015). Assessments which take place in real-world settings (e.g., performing errands in a shopping center) are time consuming and expensive to set up, lack experimental control over the external situation (e.g., Elkind et al., 2001), cannot be standardized for use in other labs, and are not feasible for certain populations (e.g., individuals with psychiatric conditions or motor difficulties; Rizzo et al., 2004; Parsons, 2015). The traditional approaches in cognitive sciences encompass the employment of static and simple stimuli, which lack ecological validity. Instead, immersive virtual reality (VR) technology enables cognitive scientists to accumulate advanced cognitive and behavioral data through the employment of dynamic stimuli and interactions with a high degree of control within an ecologically valid environment (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015). Furthermore, VR can be combined with non-invasive imaging techniques (Makeig et al., 2009; Bohil et al., 2011; Parsons, 2015), wearable mobile brain/body imaging (Makeig et al., 2009), and can be used for rehabilitation and treatment purposes (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015).

VR has great potential as an effective telemedicine tool that may resolve the current methodological problem of ecological validity (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015; Parsons et al., 2018). However, the appropriateness of VR, especially for head-mounted display (HMD) systems, is still controversial (Bohil et al., 2011; de França and Soares, 2017; Palmisano et al., 2017). The principal concern is the adverse symptomatology (i.e., nausea, dizziness, disorientation, fatigue, and instability) which stems from the implementation of VR systems (Bohil et al., 2011; de França and Soares, 2017; Palmisano et al., 2017). These adverse VR induced symptoms and effects (VRISE) endanger the health and safety of the users (Parsons et al., 2018), decrease reaction times and overall cognitive performance (Nalivaiko et al., 2015), while increasing body temperature and heart rates (Nalivaiko et al., 2015), cerebral blood flow and oxyhemoglobin concentration (Gavgani et al., 2018), brain activity (Arafat et al., 2018), and the connectivity between brain regions (Toschi et al., 2017). Hence, VRISE may compromise the reliability of cognitive, physiological, and neuroimaging data (Kourtesis et al., 2019a).

However, VRISE predominantly stem from hardware and software inadequacies, which more contemporary commercial VR hardware and software do not share (Kourtesis et al., 2019a,b). The employment of modern VR HMDs analogous to or more cutting-edge than the HTC Vive and/or Oculus Rift, in combination with ergonomic VR software, appear to significantly mitigate the presence of VRISE (Kourtesis et al., 2019a,b). However, the selection of suitable VR hardware and/or software demands acceptable technological competence (Kourtesis et al., 2019a). Minimum hardware and software features have been suggested to appraise the suitability of VR hardware and software (Kourtesis et al., 2019a). The technical specifications of the computer and VR HMD are adequate to assess their quality (Kourtesis et al., 2019a), while the virtual reality neuroscience questionnaire (VRNQ) facilitates the quantitative evaluation of software attributes and the intensity of VRISE (Kourtesis et al., 2019b).

Another limitation is that the implementation of VR technology may necessitate high financial costs, which hinders its widespread adoption by cognitive scientists. In the 90s, the cost of a VR lab with basic features cost between $20,000 and 50,000, where nowadays the cost has decreased considerably (Slater, 2018). At present, the cost of a VR lab with basic features (e.g., a HMD, external hardware, and laptop) is between

$2,000 and 2,500. However, the development of VR software is predominantly dependent on third parties (e.g., freelancers or companies) with programming and software development skills (Slater, 2018). A solution that will promote the adoption of immersive VR as a research and clinical tool might be the in-house development of VR research/clinical software by computer science literate cognitive scientists or research software engineers.

The current study endeavors to offer guidelines on the development of VR software by presenting the development of the Virtual Reality Everyday Assessment Lab (VR-EAL). Since the assessment of prospective memory, episodic memory, executive functions, and attention are likely to benefit from ecologically valid approaches to assessment, VR-EAL attempts to be one of the first neuropsychological batteries to apply immersive VR to assess these cognitive functions. However, the ecologically valid assessment of these cognitive functions demands the development of a realistic scenario with several scenes and complex interactions while avoiding intense VRISE factors.

The VR-EAL development process is presented systematically, aligned with the steps that cognitive scientists should follow to achieve their aim of designing VR studies. Firstly, the preparation stages are described and discussed. Secondly, the structure of the application (e.g., order of the scenes) is presented and discussed in terms of offering comprehensive tutorials, delivering a realistic storyline, and incorporating a scoring system. Thirdly, a pilot study is conducted to evaluate the suitability of the different versions of VR-EAL (i.e., alpha, beta, final) for implementation in terms of user experience, game mechanics, in-game assistance, and VRISE.

DEVELOPMENT OF VR-EAL

Rationale and Preparation

Prospective memory encompasses the ability to remember to initiate an action in the future (Anderson et al., 2017). The prospective memory action may be related to a specific event (e.g., when you see this person, give him a particular object) or time (e.g., at 5 p.m. perform a particular task). Attentional control processes, executive functioning, the difficulty of the filler/distractor tasks, the length of the delay between encoding the intention to perform a task and the presentation of the stimulus-cue, as well as the length of the ongoing task, all affect prospective memory ability (Anderson et al., 2017). Therefore, the VR-EAL scenarios need to incorporate both types of prospective memory actions and consider the length and difficulty of the distractor tasks and delays, as well as attentional and executive functioning.

The main theoretical frameworks of prospective memory are the preparatory attentional and memory (PAM) and the multiprocess (MP) theories (Anderson et al., 2017). The PAM theory suggests that performing prospective memory tasks efficiently requires a constant top-down monitoring for environmental and internal cues in order to recall the intended action and perform it (Smith, 2003; Smith et al., 2007). For example, an individual wants to buy a pint of milk after work. On her way home, she is vigilant (i.e., monitoring) about recognizing prompts (e.g., the sign of a supermarket) that will remind her of her intention to buy a pint of milk. In addition to PAM's top-down monitoring, MP theory suggests that bottom-up spontaneous retrieval also enables effective performance on prospective memory tasks (McDaniel and Einstein, 2000, 2007). Going back to the previous example, when the individual is not being vigilant (i.e., passive), she sees an advert pertaining to dairy products, which triggers the retrieval of her intention to buy a pint of milk. VR-EAL is required to incorporate both predominant retrieval strategies in line with these main theoretical frameworks of prospective memory (i.e., PAM and MP). This may be achieved by including scenes where the user should be vigilant (i.e., PAM) so they recognize a stimulus associated with the prospective memory task (e.g., notice a medicine on the kitchen's table in order to take it after having breakfast), as well as scenes where the user passively (i.e., MP) will attend to an obvious stimulus related to the prospective task (e.g., while being in front of the library, the user needs to remember to return a book).

Notably, the ecologically valid assessment of executive (i.e., planning and multitasking), attentional (i.e., selective visual, visuospatial, and auditory attention), and episodic memory processes is an equally important aim of VR-EAL. The relevant literature postulates that the everyday functioning of humans is dependent on cognitive abilities, such as attention, episodic memory, prospective memory, and executive functions (Higginson et al., 2000; Chaytor and Schmitter-Edgecombe, 2003; Phillips et al., 2008; Rosenberg, 2015; Mlinac and Feng, 2016; Haines et al., 2019). However, the assessment of these cognitive functions requires an ecologically valid approach to indicate the quality of the everyday functioning of the individual in the real world (Higginson et al., 2000; Chaytor and Schmitter-Edgecombe, 2003; Phillips et al., 2008; Rosenberg, 2015; Mlinac and Feng, 2016; Haines et al., 2019). However, the assessment (i.e., tasks) of these cognitive functions in VR-EAL will also serve as distractor tasks for the prospective memory components of the paradigm. Hence, the VR-EAL distractor tasks are vital to the prospective memory tasks, but at the same time, they are adequately challenging within a continuous storyline (see **Table 1**).

Furthermore, ecologically valid tasks performed in VR environments demand various game mechanics and controls to facilitate ergonomic and naturalistic interactions, and these need to be learnt by users. The scenario should include tutorials that allow users to spend adequate time learning how to navigate, use and grab items, and how the VE reacts to his/her actions (Gromala et al., 2016; Jerald et al., 2017; Brade et al., 2018; see **Table 1**). Additionally, the scenario should consider the in-game instructions and prompts offered to users such as directional arrows, non-player characters (NPC), signs, labels, ambient sounds, audio, and videos that aid performance (Gromala et al., 2016; Jerald et al., 2017; Brade et al., 2018). Importantly, this user-centered approach appears to particularly favor non-gamers in terms of performing better and enjoying the VR experience (Zaidi et al., 2018). Thus, the development of VR-EAL should be aligned with these aforementioned suggestions.

The first step of the development process was to select the target platform. In VR's case, this is the VR HMD, which allows

TABLE 1 | VR-EAL Scenario.

Order	Type	Description
Scene 1	*Tutorial*	Basic interactions and navigation
Scene 2	*Tutorial*	Interactive boards (recognition and planning)
Scene 3	Storyline	List of prospective memory tasks, shopping list (immediate recognition), and itinerary (planning)
Scene 4	*Tutorial*	List of mechanics for the prospective memory tasks, prompts, and notes
Scene 5	*Tutorial*	Cooking
Scene 6	Storyline	Prepare breakfast (multi-tasking) and take medication (prospective memory, event-based, short delay)
Scene 7	*Tutorial*	Tutorial: collect items
Scene 8	Storyline	Collect items from the living-room (selective visuospatial attention) and take a chocolate pie out of the oven (prospective memory, event-based, short delay)
Scene 9	*Tutorial*	Interaction with 3D non-player characters
Scene 10	Storyline	Call Rose (prospective memory task, time-based, short delay)
Scene 11	*Tutorial*	Gaze interaction
Scene 12	Storyline	Detect posters on both sides of the road (selective visual attention)
Scene 13	*Tutorial*	Shopping, how to collect the items from the supermarket
Scene 14	Storyline	Collect the shopping list items from the supermarket (delayed recognition)
Scene 15	Storyline	Go to the bakery to collect the carrot cake (prospective memory task, time-based, medium delay)
Scene 16	Storyline	False prompt before going to the library (prospective memory task, event-based, medium delay)
Scene 17	Storyline	Return the red book to the library (prospective memory task, event-based, medium delay)
Scene 18	*Tutorial*	Auditory interaction
Scene 19	Storyline	Detect sounds from both sides of the road (selective auditory attention)
Scene 20	Storyline	False prompt before going back home (prospective memory task, time-based, long delay)
Scene 21	Storyline	When you return home, give the extra pair of keys to Alex (prospective memory task, event-based, long delay)
Scene 22	Storyline	Put away the shopping items and take the medication (prospective memory task, time-based, long delay)

various interactions to take place within a virtual environment (VE) during the neuropsychological assessment. In our previous work (Kourtesis et al., 2019a), we have highlighted a number of suggested minimum hardware and software features which appraise the suitability of VR hardware and software. Firstly, interactions with the VE should be ergonomic in order to elude or alleviate the presence of VRISE. Also, the utilization of 6 degrees of freedom (DoF) wands (i.e., controllers) facilitates ergonomic interactions and provides highly accurate motion tracking. Lastly, the two types of HMD that exceed the minimum standards and support 6DoF controllers are the HTC Vive and Oculus Rift; hence, the target HMD should have hardware characteristics equal to or greater than these high-end HMDs (Kourtesis et al., 2019a). VR-EAL is developed to be compatible with HTC Vive, HTC Vive Pro, Oculus Rift, and Oculus Rift-S.

The second step was to select which game engine (GE) should be used to develop the VR software. For the development of VR-EAL, the feasibility of acquiring the required programming and software development skills was an important criterion for the selection of the GE because the developer of VR-EAL (i.e., the corresponding author) is a cognitive scientist who did not have any background in programming or software development. The two main GEs are Unity and Unreal. Unity requires C# programming skills, while Unreal requires C++ programming skills. Learners of C#, either experienced or inexperienced programmers, appear to experience a greater learning curve than learners of C++ (Chandra, 2012). While Unity and Unreal are of equal quality (Dickson et al., 2017), Unity as a GE has been found to be more user-friendly, and easier to learn compared to Unreal (Dickson et al., 2017). Also, Unity has an extensive online community and online resources (e.g., 3D models, software development kits; SDK), and documentation (Dickson et al., 2017). For these reasons, Unity was preferred for the development of VR-EAL. However, either Unreal or Unity would have been a sensible choice since both GEs offer high quality tools and features for software development (Dickson et al., 2017).

The final step was the acquisition of skills and knowledge. A cognitive scientist with a background either in computer or psychological sciences should have knowledge of the cognitive functions to be studied, as well as, intermediate programming and software development skills pertinent to the GE. The acquisition of these skills enables the cognitive scientist to design the VR software in agreement with the capabilities of the GE and the research aims. In VR-EAL's case, its developer meticulously studied the established ecologically valid paper-and-pencil tests such as the Test of Everyday Attention (TEA; Robertson et al., 1994), the Rivermead Behavioral Memory Test—III (RBMT-III; Wilson et al., 2008), the Behavioral Assessment of the Dysexecutive Syndrome (BADS; Wilson et al., 1998), and the Cambridge Prospective Memory Test (CAMPROMPT; Wilson et al., 2005). In addition, other research and clinical software were considered. For example, the Virtual Reality Shopping Task (Canty et al., 2014), Virtual Reality Supermarket (Grewe et al., 2014), Virtual Multiple Errands Test (Rand et al., 2009), the Invisible Maze Task (Gehrke et al., 2018), and the Jansari

Assessment of Executive Function (Jansari et al., 2014) are non-immersive VR software which assess cognitive functions such as executive functions, attentional processes, spatial cognition, and prospective memory.

Finally, the developer of VR-EAL attained intermediate programming skills in C# and software development skills in Unity. This was predominantly achieved by attending online specializations and tutorials on websites such as Coursera, Udemy, CodeAcademy, SoloLearn, and EdX. Also, a developer may consider established textbooks such as the "The VR book: Human-centered design for virtual reality" (Jerald, 2015), "3D user interfaces: theory and practice" (LaViola et al., 2017), and "Understanding virtual reality: Interface, application, and design" (Sherman and Craig, 2018). To sum it up, the acquisition of these skills enabled progression to the next stage of the development of VR-EAL, which is the writing of the scenarios/scripts.

Tutorials and Mechanics

VR-EAL commences with two tutorial scenes. The first tutorial allows the user to learn how to navigate using teleportation, to hold and manipulate items (e.g., throwing them away), how to use items (e.g., pressing a button), as well as to familiarize themselves with the in-game assistance objects (e.g., a directional arrow or a sign; see **Figure 1**). The user is prompted to spend adequate time learning the basic interactions and navigation system because these game mechanics and in-game assistance methods are fundamental to most scenes in VR-EAL.

The second tutorial instructs the user how to use interactive boards (i.e., use a map or select items from a list). This tutorial is specific to the tasks that the user should perform in the subsequent storyline scene. Similarly, the remaining tutorials are specific to their subsequent scene (i.e., the actual task) in which the user is assessed. This design enables the user to perform the tasks, without providing them with an overwhelming amount of information that may confuse the user. However, the tutorial in the fourth scene is specific to the prospective memory tasks that are performed in several scenes throughout the scenario. The instructions for all prospective memory tasks (i.e., what should be performed and when) are provided during the third scene (i.e., storyline-bedroom scene), but the first prospective memory task is not performed until the sixth scene (i.e., the cooking task; see **Table 1**).

In scene 4, the user learns how to use a VR digital watch, use prospective memory items and notes (toggle on/off the menu), and follow prospective memory prompts. These game mechanics are essential to successfully perform the prospective memory tasks. The VR digital watch is the main tool for checking the time in relation to the time-based prospective memory tasks, while the prospective memory notes are crucial reminders for the time- and event-based prospective memory tasks. Subsequently, in scene 5, the user completes a tutorial where s/he learns how to use the oven and the stove as well as the snap-drop-zones to perform the cooking task. The snap-drop-zones are game objects, which are containers that the user may attach other game objects too. In scene 7, the user learns how to collect items using the snap-drop-zones attached to the left controller (see **Figure 2**).

In scene 9, the user learns how to interact with the 3D non-player characters (NPC). The user is required to talk to the NPC to initiate a conversation (i.e., detection of a sound through the mic input), and use the interactive boards to select a response, which either presents a dichotomous choice (i.e., "yes" or "no") or a list of items (see **Figure 2**). These interactions with the NPC are central to the assessment of prospective memory, and the user should effectively interact with the NPC in six scenes to successfully perform an equal number of time- and event-based prospective memory tasks.

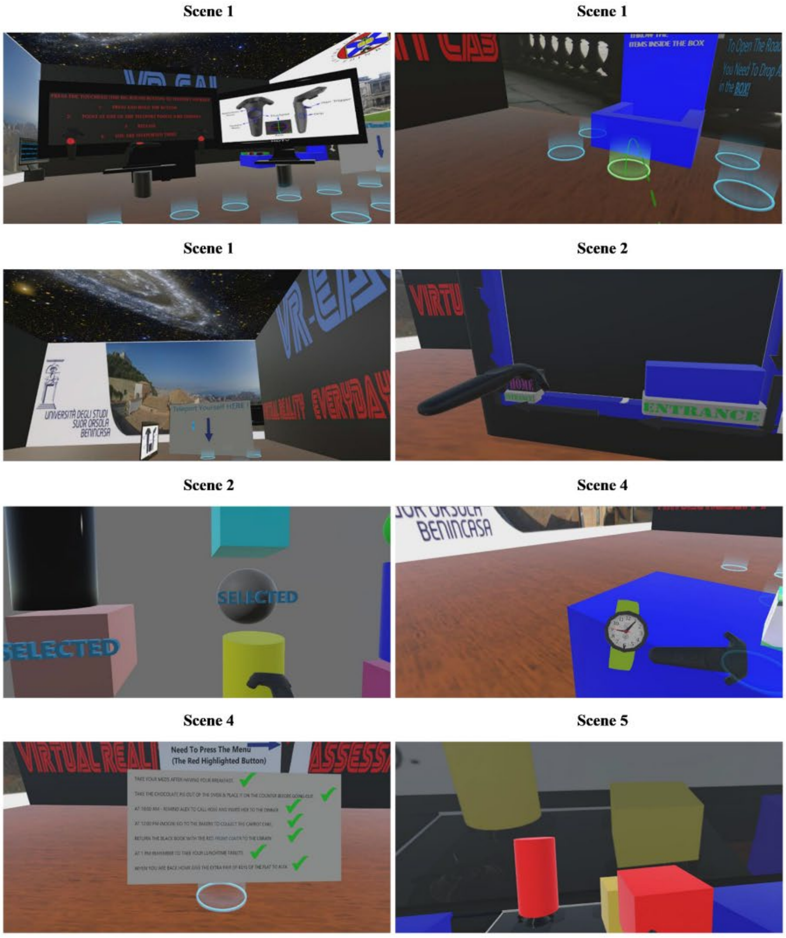
In scene 11, the user learns how to use gaze interactions. There is a circular crosshair, which indicates the collision point of a ray that is emitted from the center of the user's visual field. The user is required to direct the circular crosshair over the targets and avoid the distractors (see **Figure 2**). The user needs to effectively perform a practice trial to proceed to the next scene. The practice trial requires the user to spot the three targets and avoid all the distractors while moving. If the user is unsuccessful, then the practice trial is re-attempted. This procedure is repeated until the user effectively completes the practice trial.

Scene 13 is a short tutorial where the user is reminded how to collect items using the snap-drop-zones attached to the left controller and remove an item from the snap-drop-zone in cases where an item is erroneously picked up. In scene 18, the user learns how to detect target sounds (i.e., a bell) and avoid distractors (i.e., a high-pitched and a low-pitched bell). The user looks straight ahead and presses the trigger button on the right controller when a target sound is heard on the right side. Likewise, the user presses the trigger button on the left controller when a target sound is heard on the left side (see **Figure 2**). The sounds are activated by trigger-zones, which are placed within the itinerary of the user. This tutorial is conducted in a similar way to the scene 11 tutorial (i.e., gaze interaction). The user, while being on the move, needs to detect three target sounds and avoid the distractors to proceed to the next scene.

The time spent on each tutorial is recorded to provide the learning time for the various interaction systems (i.e., game mechanics). However, in the scene 11 and 18 tutorials, the practice trial times are also recorded. The collected data (i.e., time spent on tutorials and the attempts to complete the practice trials) for each tutorial are added to a text file that contains the user's data (i.e., performance scores on every task).

Storyline and Scoring

The required times to complete scenes and tasks are recorded. However, the task times are measured independently from the total scene times. Additionally, in the scenes where the user should perform prospective memory tasks, the number of times and the duration that the prospective memory notes appeared are also measured. These variables indicate how many times the user relies on the prospective memory notes, and how long they read them for.


Scene 1 Scene 1

Scene 1 Scene 2

Scene 2 Scene 4

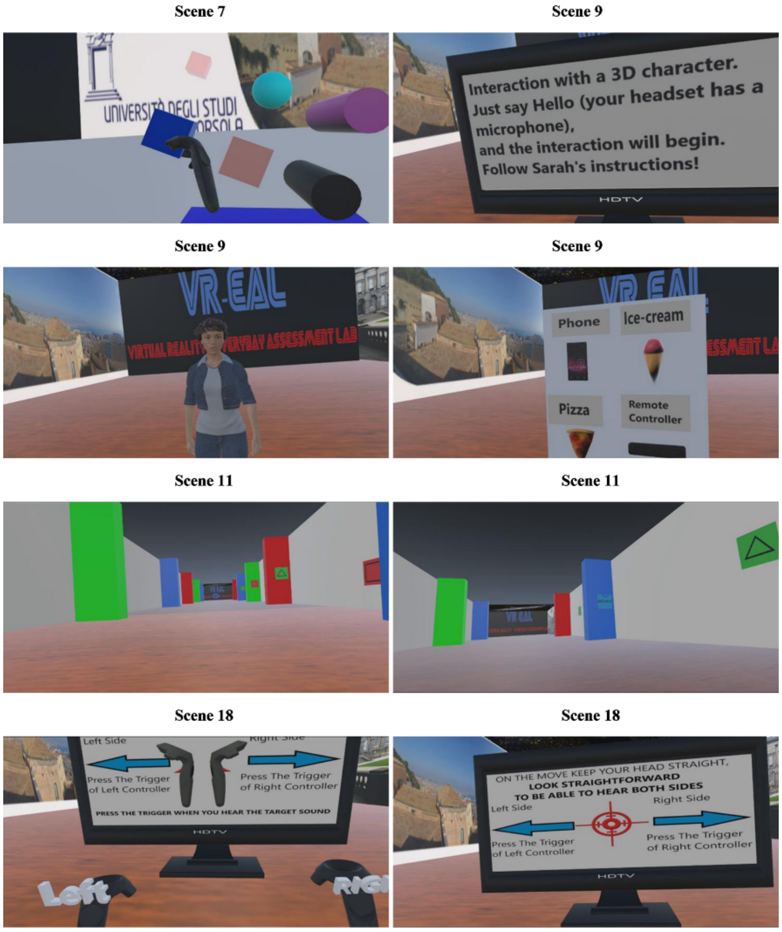
Scene 4 Scene 5

FIGURE 1 | VR-EAL tutorials: scenes 1–5.

At Home

Bedroom: immediate recognition and planning

The storyline commences in a bedroom (i.e., scene 3; see **Table 1**), where the user receives an incoming call from his/her close friend, Sarah, asking the user to carry out some errands for her (e.g., buy some shopping from the supermarket, collect a carrot cake from the bakery, return a library book). All the errands are prospective memory tasks except the shopping task. In this scene, the user should perform three different tasks. The first task is the prospective memory notes (i.e., PM-Notes) task, where the user responds affirmatively or negatively to three prompts asking the user to write down the errands (i.e., PM-tasks). The response of the user indicates his/her intention to use external tools (i.e., notes) as reminders.



Interaction with a 3D character.
Just say Hello (your headset has a microphone),
and the interaction will begin.
Follow Sarah's instructions!

Press The Trigger of Left Controller

Press The Trigger of Right Controller

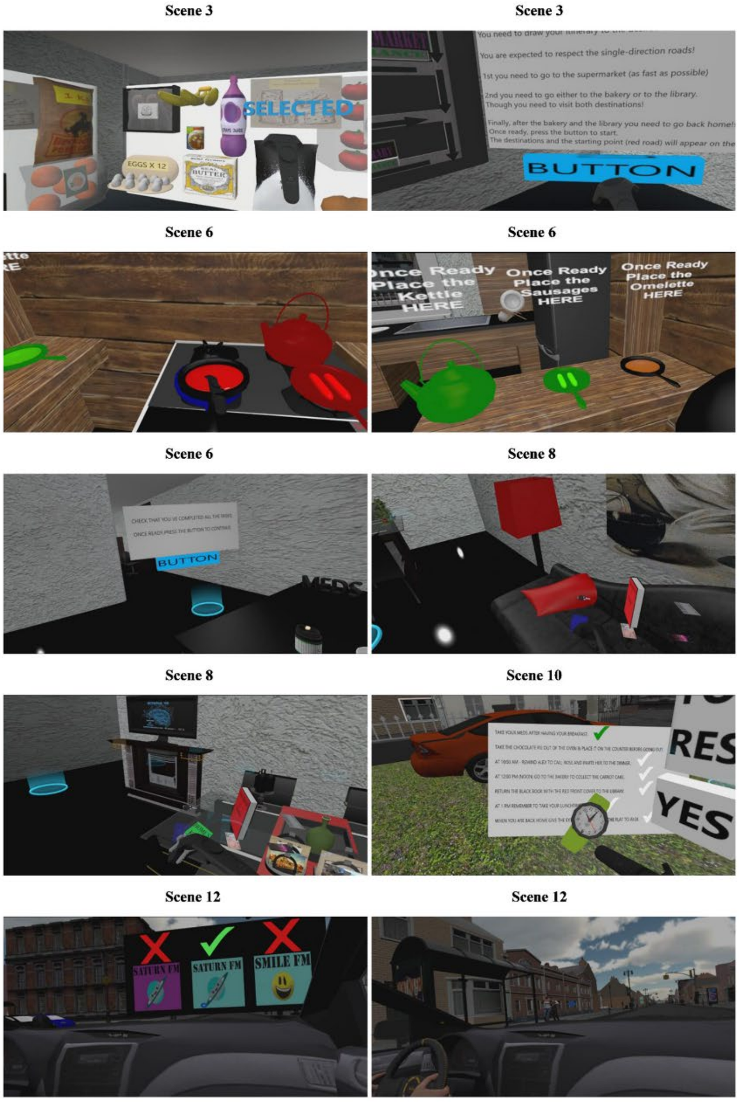
PRESS THE TRIGGER WHEN YOU HEAR THE TARGET SOUND

ON THE MOVE KEEP YOUR HEAD STRAIGHT, LOOK STRAIGHTFORWARD TO BE ABLE TO HEAR BOTH SIDES

FIGURE 2 | VR-EAL tutorials: scenes 7–18.

The second task is the immediate recognition task where the user should choose the 10 target items (i.e., create the shopping list) from an extensive array of items (see **Figure 3**), which also contains five qualitative distractors (e.g., semi-skimmed milk vs. skimmed milk), five quantitative distractors (e.g., 1 vs. 2 kg potatoes), and 10 false items (e.g., bread, bananas etc.). The user gains 2 points for each correctly chosen item, 1 point for choosing a qualitative or quantitative distractor, and 0 points for the false items. The maximum possible score is 20 points.

The third task is the planning task. The user should draw a route on a map to visit three destinations (i.e., the supermarket, bakery, and library) before returning home. The road system comprises 23 street units (see **Figure 3**). When the user selects a unit, 1 point is awarded. The ideal route to visit all three destinations is 15 units; hence, any extra or missing units are



You need to draw your itinerary

You are expected to respect the single-direction roads!

1st you need to go to the supermarket (as fast as possible)

2nd you need to go either to the bakery or to the library.
Though you need to visit both destinations!

Finally, after the bakery and the library you need to go back home!
Once ready, press the button to start.
The destinations and the starting point (red road) will appear on the

FIGURE 3 | VR-EAL storyline: scenes 3–12.

subtracted from the total possible score of 15. For example, if the user draws 18 units, then the distance from the ideal route is calculated as 3 (i.e., 18 − 15 = 3). Three is then subtracted from the ideal score of 15, resulting in a score of 12. If the user draws 12 units, the distance from the ideal route is also 3 and again 3 is subtracted from 15, resulting in a score of 12. The planning task score is also modified by the planning task completion time (e.g., a completion time two standard deviations below the average time of the normative sample is awarded 2 points while two standard deviations above the average time is subtracted 2 points).

Kitchen: multitasking and prospective memory task

In the kitchen (i.e., scene 6; see **Table 1**), the user should complete two main tasks: the cooking task (i.e., preparing breakfast) and a prospective memory task. The cooking task encompasses frying an omelet and sausages, putting a chocolate pie in the oven, as well as boiling some water for a cup of tea or coffee. The user must handle two pans (one for the omelet and one for the sausages) and a kettle. Images of the omelet and sausages are presented above the cooker to display what their appearance should be when they are ready. Scoring relies on the animations from each game object (i.e., the omelet and the sausages). At the beginning of the animation, both items have a reddish (raw) color which gradually turns to either a yellowish (omelet) or brownish (sausages) color, and finally both turn to black (burnt). The score for each pan hence depends on the time that the user removes the pans from the stove (pauses/stops the animation) and places them on the kitchen worktop (for calculation of the score, see **Figure 4**). Equally, the score for boiling the kettle is measured in relation to the stage of the audio playback (e.g., the water is ready when the kettle whistles; see **Figure 4**) when the kettle is placed on the kitchen worktop.

After breakfast, the user needs to take his/her meds (i.e., a prospective memory task). When the user has had his/her breakfast, the final button of the scene appears (see **Figure 3**). The user should press this button to confirm that all the tasks in the scene are completed. If the user has already taken his/her medication before pressing the final button, then the scene ends, and the user receives 6 points. Otherwise, the first prompt appears (i.e., "You Have to Do Something Else"). If the user then follows the prompt and takes their medication, they receive 4 points. If the user presses the final button again, then the second prompt appears (i.e., "You Have to Do Something After Having your Breakfast"). If the user follows this prompt and takes their medication, they receive 2 points. If the user presses the final button again, then the third prompt appears (i.e., "You Have to Take Your Meds"). If the user follows this prompt and takes their medication, they then receive 1 point. If the user represses the final button without ever taking their medication, they get zero points, and the scene ends.

Living room: selective visuospatial attention and prospective memory task

In the living room (i.e., scene 8; see **Table 1**), the user should collect six items (i.e., a red book, £20, a smartphone, a library card, the flat keys, and the car keys) that are placed in various locations within the living room (see **Figure 3**). The user is not required to memorize the items since there is a reminder list on one of the walls of the living room. The user collects the items by attaching them to the snap-drop-zones attached to the left controller. The user receives 1 point for each item collected. However, there are distractors (e.g., magazines, books, a remote control, a notebook, a pencil, a chessboard, and a bottle of wine) in the room. If the user attempts to collect one of the distractors, the item falls (only the target items can be attached to the snap-drop-zones), which counts as an error. After collecting all the objects, the user needs to take the chocolate pie out of the oven and place it on the kitchen worktop before leaving the apartment (prospective memory task; see scoring for medication above).

Garden: prospective memory task

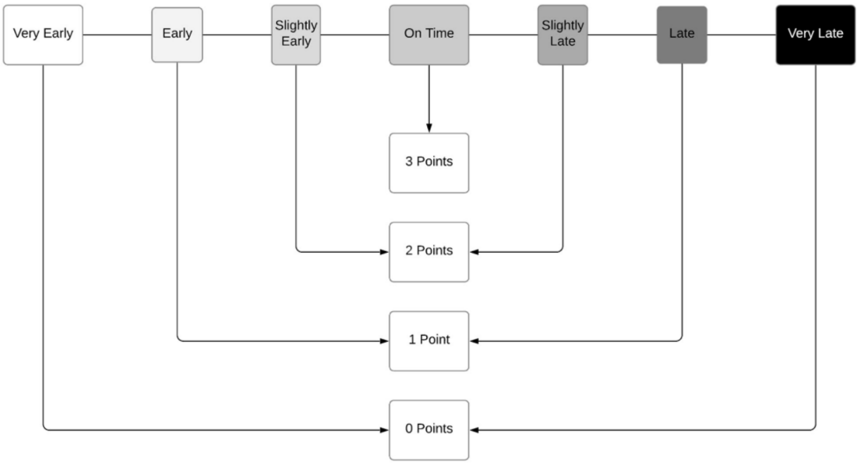
In the garden (i.e., scene 10; see **Table 1**), the user initiates a conversation with Alex (an NPC), to perform a distractor task (i.e., to open the gate). The conversation continues after this distractor task, where the user needs to respond to Alex's question (i.e., "Do we need to do something else at this time?") by selecting either "yes" or "no" (see **Figure 3**). This action is considered as the first prompt for the prospective memory task, and if the user responds "yes," then the second interactive board appears (see **Figure 5** for scoring). If the user selects "no," then the second prompt is given by Alex (i.e., "Are you sure that we do not have to do something around this time?"). If the user selects "no," then the third prompt is provided by Alex (i.e., "I think that we have to do something around this time."). If the user again selects "no," clarification is provided by Alex (i.e., "Oh yes, we need to call Rose"), and the user receives 0 points (see **Figure 5**).

When the user chooses "yes," the second interactive board appears. This second interactive board displays eight items (see **Figure 3**). There is one item, which presents the correct prospective memory response (i.e., the smartphone). There are also three items which are responses related to the other prospective memory tasks (i.e., a red book, carrot cake, flat keys). There is one item, which is semantically related to the correct prospective memory response (i.e., a tablet computer). Also, there are three items which are unrelated distractors, which are neither related to the other prospective memory tasks, nor are in the same semantic category as the correct prospective memory response (i.e., ice cream and a smartphone). Scoring depends on the user's responses on the first and the second interactive boards (see **Figure 5**).

In the City

On the road: selective visual attention

In this scenario, the user is a passenger in a car with Alex driving (i.e., scene 12; see **Table 1**). The radio is on, and the speaker announces a competition where the user needs to identify all the radio stations' target posters and avoid the distractor ones (see **Figure 3**), which are hung along the street. There are 16 target posters and 16 distractors equally allocated on both sides of the street. Eight of the distractors have the same shape as the target poster, but a different background color. The other eight distractors have the same background color as the target posters, but they are a different shape (see **Figure 3**). The user is awarded



Cooking Item	Very Early	Early	Slightly Early	On Time	Slightly Late	Late	Very Late
Pan-Omelette	0-13.99 secs	14-15.99 secs	16-17.99 secs	18-22 secs	22.01-23.99 secs	24-26 secs	> 26 secs
Pan-Sausages	0-17.99 secs	18-19.99 secs	20-21.99 secs	22-26 secs	26.01-27.99 secs	28-30 secs	> 30 secs
Kettle	0-10.99 secs	11-12.99 secs	13-14.99 secs	15-17 secs	17.01-18.99 secs	19-21 secs	> 21 secs

FIGURE 4 | A schematic representation of the cooking task scoring.

1 point when a target poster is "spotted" and subtracted 1 point when a distractor poster is "spotted." The maximum score is 16, and the number of correctly identified posters and distractors (for each type) identified on each side of the road is recorded.

Supermarket: delayed recognition and prospective memory task

The user arrives at the supermarket (i.e., scene 14; see **Table 1**), where s/he should buy the items from the shopping list. The user navigates within the shop by following the arrows, and collects the items using the snap-drop-zones attached to the left controller (see **Figure 6**). The items on the shelves of the supermarket are the same items as the immediate recognition task in scene 3 (see Bedroom). The scoring system is identical to the immediate recognition task in scene 3 (see Bedroom), and the score is calculated when the user arrives at the till to buy the items. Outside the supermarket (i.e., scene 15), the user has another conversation with Alex, where s/he needs to remember that they must collect the carrot cake at 12 noon (i.e., a prospective memory task). The conversation is performed and scored in the same way as the prospective memory task in scene 10 (see Garden). The user then goes with Alex to the bakery to collect the carrot cake.

Bakery and library: prospective memory tasks

The user is outside the bakery (i.e., scene 16; see **Table 1**), after already collecting the carrot cake. Here, they have another interaction with Alex where he asks, "Do we need to do something else at this time?" However, this time, there is no prospective memory task to perform and the user should respond negatively. This deception helps to examine whether the user is simply responding affirmatively to all prospective memory task prompts. If the user responds affirmatively (i.e., "yes"), then s/he loses points (see **Figure 7**). This conversation is similar to the prospective memory task in scene 10 (see Garden). However, the scoring is now inverted, where the user should choose "no" three times in response to Alex's prompts to avoid points being subtracted. In the prospective memory task that follows in the

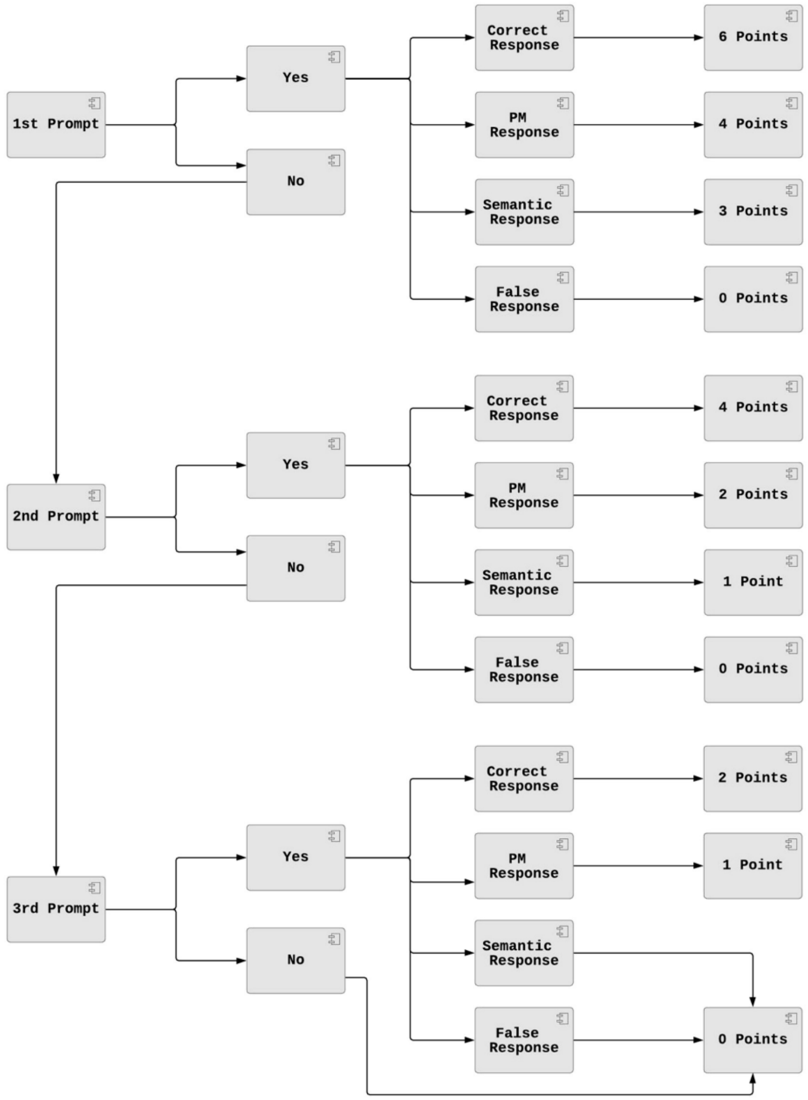


FIGURE 5 | Prospective memory: positive scoring system.

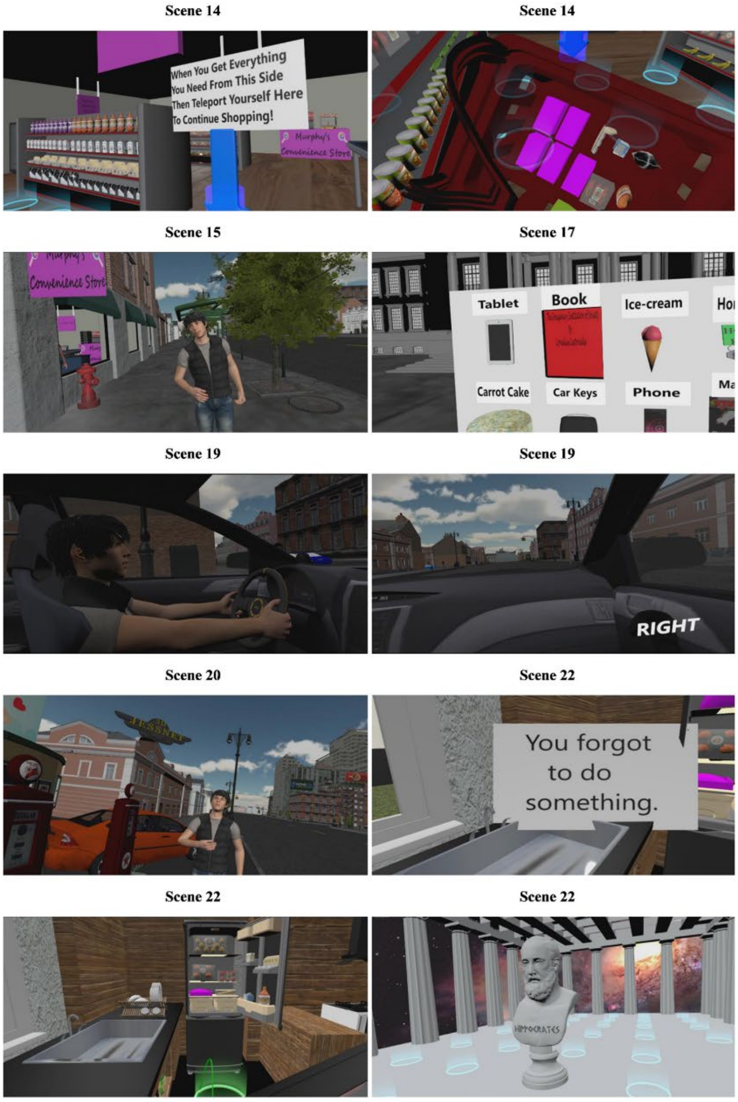

Scene 14 Scene 14
Scene 15 Scene 17
Scene 19 Scene 19
Scene 20 Scene 22
Scene 22 Scene 22

FIGURE 6 | VR-EAL storyline: scenes 14–22.

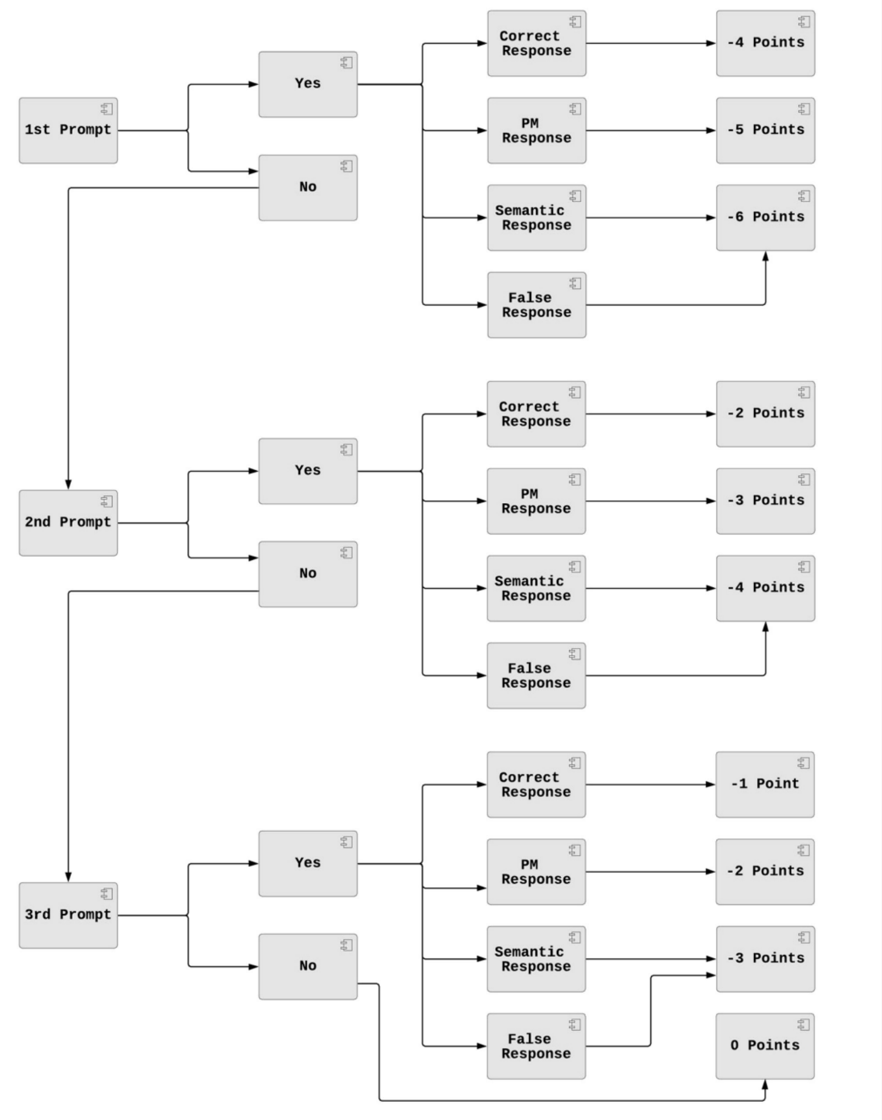


FIGURE 7 | Prospective memory: negative scoring system.

next scene, the user should again respond negatively to avoid a maximum of 3 points being deducted (see **Figure 7**). Therefore, in this task, up to 6 points may be subtracted. Then, the user arrives at the library (i.e., scene 17; see **Table 1**), where s/he has another interaction with Alex (i.e., a prospective memory task), which is performed and scored in the same way as the prospective memory task in scene 10 (see Garden and **Figure 5**). After leaving the library, Alex and the user proceed to the petrol station to refill the car.

On the road home: selective auditory attention and prospective memory tasks

The user is in the car with Alex and the radio station has another challenge (i.e., scene 19; see **Table 1** and **Figure 6**). This time small speakers playing different sounds have been placed on both sides of the street. The user should detect the target sounds and avoid the false high-pitched and low-pitched sounds while Alex drives along the street. As in the tutorial, the user presses the controller trigger when they detect a sound. If the user presses the trigger on the right controller to detect a target sound originating on the right side, then s/he gets 2 points. If the user presses the trigger on the left controller to detect a target sound originating on the left side, s/he also gets 2 points. If the user presses the trigger on the right controller to detect a target sound originating on the left side or a trigger on the left controller to detect a target sound originating on the right side, s/he gains only 1 point. On the other hand, if the user responds to a distractor sound, irrelevant of its origin or the controller used to respond, 1 point is deducted. The stored data summarize the number of detected sounds of each type (i.e., target sounds, low pitched distractor sounds, high pitched distractor sounds), the number of sounds detected on the left and right sides, and how many times the wrong controller (i.e., false side) was used to detect a target sound.

After the car ride, the user is at the petrol station with Alex (i.e., scene 20; see **Table 1**). The user has another conversation with Alex, where s/he receives false prompts (i.e., there is not a prospective memory task to perform). This prospective memory task is performed and scored in the same way as the Bakery prospective memory task (i.e., scene 20, see **Figure 7**). Then, the user returns back home with Alex (i.e., scene 21), where the user has their last interaction with Alex, and should give him the extra pair of keys to the flat. This prospective memory task is also performed and scored as the prospective memory task in scene 10 (i.e., see Garden and **Figure 5**).

Back Home: distractor and prospective memory task

In the final scene (i.e., scene 22), the user is back home, where s/he is required to perform two tasks (see **Figure 6**). The first task is a distractor task, where the user needs to put away the items that s/he has bought from the supermarket. While doing this, s/he needs to remember that s/he should take his/her medication at 1 p.m. If the user performs the task on time, then s/he receives 6 points. If the user fails to remember the prospective memory task after 70 s, a prompt appears. If the user performs the task after this first prompt, s/he receives 4 points. If the user ignores the first prompt, after another 10 s, a second prompt appears. If the user performs the task after the second prompt, s/he receives 2 points. If the user ignores the second prompt, after a further 10 s, a third and final prompt appears. If the user performs the task after the third prompt, s/he receives 1 point. If the user ignores the third prompt and presses the final button, s/he receives 0 points.

Once the user presses the final button, the scenario finishes and the credits appear. Here, the user is informed of the aims of VR-EAL. The VR-EAL attempts to be an extensive neuropsychological assessment of prospective memory, episodic memory, executive functions, and attentional processes by collecting various data pertinent to these cognitive functions (see **Supplementary Material I** for an example of VR-EAL data).

Development of VR Software in Unity

The scenario provides the main framework for developing the VR application. VR-EAL was developed to be compatible with the HTC Vive, HTC Vive Pro, Oculus Rift, and Oculus Rift-S to be aligned with the minimum hardware technological specifications for guaranteeing health and safety standards and the reliability of the data (Kourtesis et al., 2019a). The quality of VR-EAL was assessed in terms of user experience, game mechanics, in-game assistance, and VRISE using the Virtual Reality Neuroscience Questionnaire (VRNQ; Kourtesis et al., 2019b). The total duration for the VR neuropsychological assessment is ∼60 min, which falls within the suggested maximum duration range for VR sessions (Kourtesis et al., 2019b). Long VR sessions appear more susceptible to VRISE, though, long (50–70 min) VR sessions which exceed the parsimonious cut-offs from the VRNQ do not induce VRISE (Kourtesis et al., 2019b). For this reason, the parsimonious cut-offs for the VRNQ (see **Table 2**) will be used to ensure that VR-EAL users do not suffer from VRISE (Kourtesis et al., 2019b).

The development of VR-EAL should be proximal to commercial VR applications. The first step of the development is to select Unity's settings to support the development of VR software. For the development of VR-EAL, Unity version 2017.4.8f1 was used. Unity supports VR software development kits (SDK). The built-in support for the SDKs is for the OpenVR SDK and the Oculus SDK. In the player settings of Unity, the developer may select the VR/XR supported box, which allows the addition of the aforementioned SDKs. For VR-EAL, Unity's

TABLE 2 | VRNQ minimum and parsimonious cut-offs.

Score	Minimum cut-offs	Parsimonious cut-offs
User experience	≥25/35	≥30/35
Game mechanics	≥25/35	≥30/35
In-game assistance	≥25/35	≥30/35
VRISE	≥25/35	≥30/35
VRNQ Total Score	≥100/140	≥120/140

The median of each sub-score and total score should meet the suggested cut-offs to determine that the evaluated VR software is of adequate quality without any significant VRISE. The utilization of the parsimonious cut-offs more robustly supports the suitability of the VR software. Derived from Kourtesis et al. (2019b).

support for both the OpenVR SDK and the Oculus SDK were added, though, priority was given to the OpenVR SDK.

Navigation and Interactions

VR software for the cognitive sciences may require intensive movement and interactions. However, the development of such interactions demands highly advanced programming skills in C# and expertise in VR software development in Unity. Nonetheless, on Unity's asset store and GitHub's website, there are some effective alternatives that facilitate the implementation of intensive interactions without the requirement of highly advanced software development skills. The utilization of the SteamVR SDK, Oculus SDK, Virtual Reality Toolkit (VRTK) or similar toolkits and assets are options which should be considered. For the development of VR-EAL, the SteamVR SDK and VRTK were selected to develop accurate interactions compatible with the capabilities of the 6DoF controllers of HTC Vive and Oculus Rift. The advantage of SteamVR SDK, which was developed based on OpenVR SDK, is that is compatible with both the HTC Vive and Oculus Rift, though, it does not offer a wide variety of interactions or good quality physics. Nonetheless, the VRTK mounts the SteamVR SDK and offers better quality physics and plenty of interactions that support the development of VR research software for cognitive sciences.

A fundamental interaction in the VE is navigation. HTC Vive and Oculus Rift offer a play area of an acceptable size, which permits ecologically valid scenarios and interactions to be developed (Porcino et al., 2017; Borrego et al., 2018). However, the VR play area is restricted to the limits of the physical space and tracking area; hence, it does not allow navigation which is based on physically walking (Porcino et al., 2017; Borrego et al., 2018). A suitable solution is the implementation of a navigation system based on teleportation. Teleportation enables navigation exceeding the boundaries of the VR play area and delivers high-level immersion, a pleasant user experience, and decreases the frequency of VRISE. Typically, a navigation system of a VR software which depends on a touchpad, keyboard, or joystick, substantially increasing the frequency and intensity of VRISE (Bozgeyikli et al., 2016; Frommel et al., 2017; Porcino et al., 2017). In VR-EAL, a combination of teleportation and physical movement (i.e., free movement of the upper limbs and walking in a small-restricted area) is used (see **Figures 1–3**, **6**).

The VRTK provides scripts and tools that aid the developer to build a teleportation system. The VRTK is compatible with 6DoF controllers, which are necessary to provide naturalistic and ergonomic interactions. In addition, the implementation of 6DoF controllers facilitates familiarization with their controls and their utilization, because they imitate real life hand actions and movements (Sportillo et al., 2017; Figueiredo et al., 2018). The VR-EAL user learns the controls in the tutorials, though, there are also in-game instructions and aids that assist even a non-gamer user to grab, use, and manipulate items. These in-game assistance methods significantly alleviate the occurrence of VRISE, while increasing the user's level of enjoyment (Caputo et al., 2017; Porcino et al., 2017). Finally, the VRTK offers additional gamified interactions through the snap-drop-zones. The snap-drop-zones are essentially carriers of game objects and their mechanics are similar to the trigger-zones. For example, when a game object (i.e., a child object of a controller) enters the zone, if the game object is released (i.e., stops being a child object of the controller), this game object is attached to the snap-drop-zone (i.e., it becomes a child object). In VR-EAL, the snap-drop-zones are extensively used, allowing the scoring of tasks, which otherwise would be less effective in terms of accuracy of response times.

The interaction and navigation systems are essential to increase immersion. However, immersion depends on the strength of the placement, plausibility, and embodiment illusions (Slater, 2009; Slater et al., 2010; Maister et al., 2015; Pan and Hamilton, 2018). An ecologically valid neuropsychological assessment necessitates genuine responses from the user. Robust placement and plausibility illusions ensure that the user will genuinely perform the tasks as s/he would perform them in real life (Slater, 2009; Slater et al., 2010; Pan and Hamilton, 2018). The placement illusion is the deception of the user that s/he is in a real environment and not in a VE (Slater, 2009; Slater et al., 2010). However, the placement illusion is fragile because the VE should react to the user's actions (Slater, 2009; Slater et al., 2010). This is resolved by the plausibility illusion, which is the deception of the user that the environment reacts to his/her actions. Therefore, the user believes the plausibility of being in a real environment (Slater, 2009; Slater et al., 2010). The naturalistic interactions in the VE that VRTK and SteamVR SDK offer are pertinent to the plausibility illusion.

Graphics

A strong placement illusion relies on the quality of the graphics and 3D objects (Slater, 2009; Slater et al., 2010). Correspondingly, the quality of the graphics principally depends on the rendering (Lavoué and Mantiuk, 2015). The rendering comprises the in-game quality of the image (i.e., perceptual quality), and the omission of unnecessary visual information (i.e., occlusion culling) (Lavoué and Mantiuk, 2015). The advancement of these rendering aspects ameliorates both the quality of graphics and the performance of the VR software (Brennesholtz, 2018). Likewise, the amplified image refresh rate and resolution decrease the frequency and intensity of VRISE (Brennesholtz, 2018). However, the rendering pipeline and shaders in Unity are not optimized to meet VR standards. The VR software developer should select different rendering options, so the quality of graphics is good and the image's refresh rate is equal to or above 90 Hz, which is the minimum for high-end HMDs like the HTC Vive and Oculus Rift. For example, the "Lab renderer" is an asset that allows VR optimized rendering and replaces the common shaders with VR optimized ones. Additionally, the "Lab renderer" supports an extensive number of light sources (i.e., up to 15), which otherwise would not be feasible in VR. However, the developer needs to build a global illumination map (i.e., lightmap), which substantially alleviates the cost of lights and shadows on the software's performance (Jerald, 2015; LaViola et al., 2017; Sherman and Craig, 2018). Usually, the lightmapping process is the final step in the development process.

The acquisition of 3D objects may be expensive or time-consuming. However, there are several free 3D objects on Unity's asset store and webpages, such as TurboSquid and Cgtrader, which can be used for the development of VR research software. Importantly, the license for these 3D objects obliges the developer not to use them for commercial purposes. However, research VR software like VR-EAL is free, and research software developers usually do not commercialize their products. Although there are several free 3D objects on the websites mentioned above, it is likely that these 3D objects are not compatible with VR standards. In VR, the 3D objects should comprise a low number of polygons (Jerald, 2015; LaViola et al., 2017; Sherman and Craig, 2018). A decrease in polygons may be achieved using software like 3DS Max. The optimization of the 3D objects (to meet VR standards) may be achieved by simply importing the 3D objects, optimizing them, and then exporting them with a low number of polygons in a Unity compatible format (i.e., fbx and obj).

Nevertheless, developers often aim to create large VEs such as cities, towns, shops, and neighborhoods. Each 3D object, whether it be small (e.g., a pen), medium (e.g., a chair), or large (e.g., a building), may comprise several mesh renderers. Unity requires one batch (i.e., draw call) for each mesh renderer. In large environments, the batching may significantly lower the image's refresh rate and the overall performance of the software (Jerald, 2015; LaViola et al., 2017; Sherman and Craig, 2018). However, assets like MeshBaker are designed to solve this problem. MeshBaker merges all the selected textures and meshes into a clone game object with a small number of meshes and textures. For example, the town that was designed for VR-EAL required >1,000 draw calls. After the implementation of MeshBaker, the draw calls were decreased to 16. However, the disadvantage of MeshBaker is that it does not clone the colliders. Hence, the developer needs to deactivate the mesh renderers of the original game object(s) and leave active all the colliders, while the original game object(s) should be precisely in the same position with the clone(s) so the colliders of the former are aligned with the meshes of the latter. Of note, MeshBaker should be purchased from Unity's asset store in contrast with the other assets used in VR-EAL's development which are freely available (i.e., SteamVR SDK, VRTK, and Lab renderer).

Sound

Another important aspect of VR software development is the quality of the sound. The addition of spatialized sounds in the VE (e.g., ambient and feedback sounds) augments the level of immersion and enjoyment (Vorländer and Shinn-Cunningham, 2014), and significantly reduces the frequency of VRISE (Viirre et al., 2014). Spatialized sounds in VR assist the user to orient and navigate (Rumiński, 2015), and enhance the geometry of the VE without reducing the software's performance (Kobayashi et al., 2015). In Unity, a developer may use tools like SteamAudio, Oculus Audio Spatializer, or Microsoft Audio Spatializer for good quality and spatialization of the audio aspects. In VR-EAL's development, Steam Audio was used. SteamAudio spatializes the sound to the location of the audio source's location and improves the reverberance of sounds (i.e., Unity's reverb zone). Notably, the strength of the plausibility illusion is analogous to the sensorimotor contingency, which is the integration of the senses (i.e., motion, vision, touch, smell, taste) (Gonzalez-Franco and Lanier, 2017). Moreover, the VRTK enables the utilization of a haptic modality. For example, when the user grabs an item in the VE, s/he expects a tactile sense as would be experienced in real life. The haptic feedback of the VRTK allows the developer to activate/deactivate the vibration system of the 6 DoF controllers when an event occurs (e.g., grabbing or releasing a game object) and define the strength and the duration of the vibration. The spatialized audio and the haptics additionally reinforce the plausibility illusion by providing an expected auditory and haptic feedback to the user (Jerald, 2015; LaViola et al., 2017; Sherman and Craig, 2018).

3D Characters

Furthermore, VR research software like VR-EAL, which includes social interactions with virtual characters, should also consider the quality of the 3D characters in terms of realistic appearance and behavior. For example, Morph 3D and Mixamo both offer free and low-cost realistic 3D characters that may be used in VR software development. For VR-EAL, Morph 3D was preferred, though, other virtual humans from Unity's asset store were used to populate the scenes (e.g., individuals waiting for the bus at the bus stop). The 3D characters provided by Morph 3D have modifiable features, which may be used by the developer to customize the character's appearance (e.g., body size) and expressions (e.g., facial expressions which signify emotions such as happiness and sadness). Morph 3D provides two free 3D characters (i.e., female and male) capable of displaying naturalistic behavior (i.e., body and facial animations). The developer may use body animations which derive from motion capture (MoCap) techniques. For the development of VR-EAL, body animations were derived from free sample animations from Unity's Asset Store (e.g., hand movement during talking, and waving) and the MoCap animations library of the Carnegie Mellon University. However, the effective implementation of the animations requires modification and synchronization (e.g., the animation should be adjusted to the length of the 3D character's interaction) using Unity's animation and the animator's windows. The animation window may be used for synchronization, while the animator is a state machine controller that controls the transition between animations (e.g., when this event happens, play this animation, or when animation X ends, play animation Y).

However, the most challenging aspect of realistic 3D characters is the animation of their facial features. The 3D character should have realistic eye interactions (i.e., blinking, looking at or away from the user) and talking (i.e., a realistic voice and synchronized lip movements). Limitations in both time and resources did not allow for seamless face and body animations since that would require multimillion dollars' worth of equipment like those used by big game studios. This limitation can result in an uncanny valley effect (Seyama and Nagayama, 2007; Mori et al., 2012). However, previous research has shown that, when users interact with 3D humanoid embodied agents that have the role of an instructor (like the ones used in VR-EAL), they have less expectations for that character due to their role and limited interactivity (Korre, 2019). The addition of 3D characters was important because they deliver an interaction metaphor resembling human-to-human interactions (Korre, 2019). Even though adding a 3D character in the scene can introduce biases, the illusion of humanness—which is defined as the user's notion that the system (in this case the 3D NPC) possesses human attributes and/or cognitive functions—has been found to increase usability (Korre, 2019).

Realistic voices may be established by employing voice actors to produce the script. However, the employment of temporary staff increases development costs. For VR-EAL, text-to-speech technologies were used as an alternative solution to deliver realistic voices. Balabolka software was used in conjunction with Ivona3D Voices (n.b., Ivona3D has been replaced by Amazon Polly). Balabolka is an IDE for text-to-speech which allows further manipulation of voices (i.e., pitch, rate, and volume), while Ivona3D provides realistic voices. The developer types or pastes the text into Balabolka, Balabolka modifies it with respect to the desired outcome (e.g., high-pitched or low-pitched voice) and exports the file in a.wav format. Additionally, free software like Audacity may be used, which offers greater variety in sound modifications. The second crucial part is to synchronize the eyes and lip movements with the voice clips and body animations. There are assets on Unity's asset store that may be used to achieve this desired outcome. In VR-EAL, Salsa3D and RandomEyes3D were used to attain good quality facial animations and lip synchronization. Salsa3D synchronizes the lips with the voice clip, while RandomEyes3D allows the developer to control the proportion of eye contact with the user for each voice clip.

Summary of the VR-EAL Illusions

Summing up, the described VR-EAL development process facilitates the utilization of ergonomic interactions, a VR compatible navigation system, good quality graphics, haptics, and sound, as well as social interactions with realistic 3D characters. These software features contribute to the lessening or avoidance of VRISE and augmentation of the level of immersion by providing placement and plausibility illusions. However, VR-EAL does not seem to deliver a strong embodiment illusion (i.e., the deception that the user owns the body of the virtual avatar), because it only relies on the presence of the 6 DoF controllers. A possible solution would be the implementation of inverse kinematics, which animates the virtual avatar with respect to the user's movements. In addition, the temporal illusion (i.e., deceiving the user into thinking that the virtual time is real-time) only relies on changes in environmental cues (e.g., the movement of the sun, and changes in lighting). Therefore, a VR digital watch was developed (freely distributed on GitHub) and used in an attempt to increase the strength of the temporal illusion. To conclude, the development of VR research software is feasible mainly using free or low-cost assets from GitHub, Unity Asset's store, and other webpages. However, the suitability and quality of the VR software should be evaluated before its implementation in research settings.

EVALUATION OF VR-EAL

Participants

Twenty-five participants (six female gamers, six male gamers, seven female non-gamers, and six male non-gamers) were recruited for the study via the internal email network of University of Edinburgh as well as social media. The mean age of the participants was 30.80 years (SD = 5.56, range = 20–45) and the mean years of full-time education was 14.20 years (SD = 1.60, range = 12–16). Twelve participants (three female gamers, three male gamers, three female non-gamers, and three male non-gamers; mean age = 30.67 years, SD = 2.87, range = 26–36; mean educational level = 14.75 years, SD = 1.30, range = 12–16 years) attended all three VR sessions (i.e., alpha, beta, and final versions), while the remaining 13 participants only attended the final version session. The gamer experience was a dichotomous variable (i.e., gamer or non-gamer) based on the participants' response to a question asking whether they played games on a weekly basis. The current study has been approved by the Philosophy, Psychology and Language Sciences Research Ethics Committee of the University of Edinburgh. All participants were informed about the procedures, possible adverse effects (e.g., VRISE), data utilization, and the general aims of the study both orally and in writing; subsequently, every participant gave written informed consent.

Material

Hardware and Software

An HTC Vive HMD, two lighthouse-stations for motion tracking, and two 6 DoF controllers were used. The HMD was connected to a laptop with a 2.80 GHz Intel Core i7 7700HQ processor, 16 GB RAM, a 4,095 MB NVIDIA GeForce GTX 1070 graphics card, a 931 GB TOSHIBA MQ01ABD100 (SATA) hard disk, and Realtek High Definition Audio. The size of the VR play area was 4.4 m^2. The software was the alpha version of VR-EAL for session 1, the beta version of VR-EAL for session 2, and the final version of VR-EAL for session 3.

VRNQ

The VRNQ is a paper-and-pencil questionnaire containing 20 questions, where each question refers to one of the criteria necessary to assess VR research/clinical software in neuroscience (Kourtesis et al., 2019b). The 20 questions assess four domains: user experience, game mechanics, in-game assistance, and VRISE. The VRNQ has a maximum total score of 140, and 35 for each domain. VRNQ responses are indicated on a 7-point Likert style scale ranging from 1= extremely low to 7 = extremely high. Higher scores indicate a more positive outcome; this also applies to the evaluation of VRISE intensity. Hence, higher VRISE scores indicate lower intensities of VRISE (i.e., 1 = extremely intense feeling, 2 = very intense feeling, 3 = intense feeling, 4 = moderate feeling, 5 = mild feeling, 6 = very mild feeling, 7 = absent). Additionally, the VRNQ allows participants to provide qualitative feedback, which may be useful during the development process. Lastly, the VRNQ has two cut-off scores, the minimum (i.e., 25 for every sub-score, and 100 for the total score) and parsimonious (i.e., 30 for every sub-score, and 120 for the total score) cut-offs. The median scores derived from the user sample should exceed at least the minimum cut-offs, while for VR software which requires long VR sessions, then the parsimonious cut-offs should be preferred. For the evaluation of VR-EAL, the parsimonious cut-offs were opted to support the suitability of VR-EAL. The VRNQ can be downloaded from **Supplementary Material II.**

Procedures

Twelve participants attended all three VR sessions, while an additional 13 participants only attended the third session. The period between each session was 6–8 weeks. In each session, participants were immersed in a different version of VR-EAL. Each session began with inductions in VR-EAL, the HTC Vive, and the 6 DoF controller. Then, participants played a version of VR-EAL. Lastly, after the completion of VR-EAL, participants were asked to complete the VRNQ. A preview of the final version of VR-EAL can be found in **Supplementary Material III** or by following the hyperlink: https://www.youtube.com/watch?v=IHEIvS37Xy8andt=.

Statistical Analysis

Bayesian statistics were preferred over null hypothesis significance testing (NHST). *P*-values calculate the distance (i.e., the difference) between the data and the null hypothesis (H0) (Cox and Donnelly, 2011; Held and Ott, 2018). The *p*-values assess the assumption of no difference or no effect, while the Bayesian factor (BF_{10}) converts *p*-values into evidence in favor of the alternative hypothesis (H1) against the H0 (Cox and Donnelly, 2011; Held and Ott, 2018). BF_{10} is found robustly more parsimonious than the *p*-value in evaluating the evidence against the H0 (Cox and Donnelly, 2011; Held and Ott, 2018; Wagenmakers et al., 2018a,b). Importantly, the difference between BF_{10} and *p*-values is even greater (in favor of BF_{10}) in small sample sizes, where BF_{10} should be opted for as it is more parsimonious (Held and Ott, 2018; Wagenmakers et al., 2018a,b). For these reasons, the BF_{10} was preferred instead of *p*-values for the assessment of statistical inference, especially while having a relatively small sample size. Moreover, a larger BF_{10} postulates more evidence in support of H1 (Cox and Donnelly, 2011; Marsman and Wagenmakers, 2017; Held and Ott, 2018; Wagenmakers et al., 2018a,b). Specifically, a $BF_{10} \leq 1$ indicates no evidence in favor of H1, while $1 < BF_{10} < 3$ indicates anecdotal evidence for H1, $3 \leq BF_{10} < 10$ indicates moderate evidence for H1, $10 \leq BF_{10} < 30$ indicates strong evidence for H1, $30 \leq BF_{10} < 100$ indicates very strong evidence for H1, and a $BF_{10} \geq 100$ indicates extreme evidence for H1 (Marsman and Wagenmakers, 2017; Wagenmakers et al., 2018a,b). For our analyses, we accept the notion put forward by Marsman and Wagenmakers (2017), Wagenmakers et al. (2018a,b) of $BF_{10} \leq 1$ indicating no evidence in favor of H1, $BF_{10} > 3$ indicating moderate evidence in favor of H1, $BF_{10} \geq 10$ indicating strong evidence for H1, and $BF_{10} \geq 100$ indicating extreme evidence for H1. In this study, a parsimonious threshold of $BF_{10} \geq 10$ was set for statistical inference, which postulates strong evidence in favor of the H1 (Marsman and Wagenmakers, 2017; Wagenmakers et al., 2018a,b), and corresponds to a $p < 0.01$ (e.g., $BF_{10} = 10$) or to a $p < 0.001$ (e.g., $BF_{10} > 11$) (Cox and Donnelly, 2011; Held and Ott, 2018). However, we report both BF_{10} and *p*-values in this study. A Bayesian paired samples *t*-test was performed to compare the VRNQ results for each version of VR-EAL ($N = 12$), as well as to inspect potential differences between gamers ($N = 12$) and non-gamers ($N = 13$). The Bayesian statistical analyses were performed using JASP (Version 0.8.1.2) (JASP Team, 2017).

Results

There was not a significant difference between gamers and non-gamers in VRNQ scores (see **Table 3**). The final version of VR-EAL exceeded the parsimonious cut-off for the VRNQ total score, while the alpha and beta versions of VR-EAL did not (see **Table 4**). Notably, the VRNQ sub-scores of the final version of VR-EAL also exceeded the parsimonious VRNQ cut-offs (see **Table 4**), while the average duration of the VR sessions (i.e., duration of being immersed) was 62.2 min (SD = 5.59) across the 25 participants. The beta version of VR-EAL approached the cut-offs for user experience and game mechanics; however, it was substantially below the cut-offs for in-game assistance and VRISE. The alpha version of VR-EAL was significantly below the cut-offs for every sub-score of VRNQ.

According to the adopted nomenclature (i.e., $BF_{10} \leq 1$ indicating no evidence in favor of H1, $BF_{10} > 3$ indicating moderate evidence in favor of H1, $BF_{10} \geq 10$ for H1, and $BF_{10} \geq 100$ indicating extreme evidence for H1) by Marsman and Wagenmakers (2017) and Wagenmakers et al. (2018a,b), the Bayesian *t*-test analysis ($N = 12$) demonstrated significant differences in the VRNQ scores between the final, beta, and alpha versions of the VR-EAL (see **Table 5**). We observed that the probability of the alternative hypothesis that the VRNQ total score for the final version is greater than the VRNQ total score for the alpha version is 57,794 times greater (i.e., $BF_{10} = 57{,}974$; see **Table 5**) than the probability of H0 (i.e., not being greater). Similarly, the probability of the alternative hypothesis that the VRNQ total score for the final version is greater than the VRNQ total score for the beta version is 855 times greater (i.e., $BF_{10} = 855$; see **Table 5**) than the probability of H0. Lastly, the probability of the alternative hypothesis that the VRNQ total score for the beta version is greater than the VRNQ total score for the alpha version is 101 times greater (i.e., $BF_{10} = 101$; see **Table 5**) than the probability of H0. The remaining alternative hypotheses for the comparisons between the versions of VR-EAL

TABLE 3 | Comparison of VRNQ scores between gamers and non-gamers.

VRNQ scores	*p*-value	BF_{10}	Error %
Total VRNQ	$p = 0.631$	0.402	1.052e−4
User experience	$p = 0.289$	0.546	0.001
Game mechanics	$p = 0.459$	0.429	2.003e−4
In-game assistance	$p = 0.841$	0.374	0.030
VRISE	$p = 0.983$	0.368	0.030

$BF_{10} > 10$; **$BF_{10} > 30$; *$BF_{10} > 100$; No significant differences observed.*

TABLE 4 | VRNQ scores for alpha, beta, and final version of VR-EAL.

	N	Median (MAD)	Cut-off	Maximum score
Total VRNQ—alpha version	12	100 (6)	$\geq$120	140
User experience—alpha version	12	25 (2)	$\geq$30	35
Game mechanics—alpha version	12	23.5 (3.5)	$\geq$30	35
In-game assistance—alpha version	12	24 (3)	$\geq$30	35
VRISE—alpha version	12	25.5 (1.5)	$\geq$30	35
Total VRNQ—beta version	12	109.5 (2.5)	$\geq$120	140
User experience—beta version	12	28 (1)	$\geq$30	35
Game mechanics—beta version	12	29 (1)	$\geq$30	35
In-game assistance—beta version	12	26 (1)	$\geq$30	35
VRISE—beta version	12	26 (1)	$\geq$30	35
Total VRNQ—final version—all	25	128 (5)	$\geq$120	140
User experience—final version—all	25	31 (2)	$\geq$30	35
Game mechanics—final version—all	25	32 (2)	$\geq$30	35
In-game assistance—final version—all	25	32 (3)	$\geq$30	35
VRISE—final version—all	25	33 (1)	$\geq$30	35
Total VRNQ—final version—gamers	12	129.5 (5)	$\geq$120	140
User experience—final version—gamers	12	32.5 (1.5)	$\geq$30	35
Game mechanics—final version—gamers	12	32 (1.5)	$\geq$30	35
In-game assistance—final version—gamers	12	32.5 (2)	$\geq$30	35
VRISE—final version—gamers	12	33 (1)	$\geq$30	35
Total VRNQ—final version—non-gamers	13	128 (4)	$\geq$120	140
User experience—final version—non-gamers	13	31 (1)	$\geq$30	35
Game mechanics—final version—non-gamers	13	31 (2)	$\geq$30	35
In-game assistance—final version—non-gamers	13	32 (3)	$\geq$30	35
VRISE—final version—non-gamers	13	33 (2)	$\geq$30	35

MAD, Median Absolute Deviation.

TABLE 5 | Bayesian paired sample *t*-test results.

Alternative Hypothesis (H1)	*p*-value	BF$_{10}$	Error %
Total VRNQ—alpha < Total VRNQ—beta	$p < 0.001$	101.651***	$\sim$ 2.226e−5
Total VRNQ-alpha < Total VRNQ-final	$p < 0.001$	57974.267***	$\sim$ 9.361e-35
Total VRNQ-beta < Total VRNQ-final	$p < 0.001$	855.603***	$\sim$ 1.506e-17
User experience-alpha < User experience-beta	$p < 0.001$	21.221*	$\sim$ 9.875e−5
User experience-alpha < User experience-final	$p < 0.001$	681.518***	$\sim$ 8.429e-24
User experience-beta < User experience-final	$p < 0.001$	17.597*	$\sim$ 2.172e−4
Game mechanics-alpha < Game mechanics-beta	$p < 0.001$	47.214**	$\sim$ 1.820e−4
Game mechanics-alpha<Game mechanics-final	$p < 0.001$	487.798***	$\sim$ 2.337e-19
Game mechanics-beta <Game mechanics-final	$p < 0.001$	17.262*	$\sim$ 2.288e−4
In-game assistance-alpha < In-game assistance-beta	$p = 0.098$	1.095	$\sim$ 9.459e−4
In-game assistance-alpha <In-game assistance-final	$p < 0.001$	224.329***	$\sim$ 1.110e-18
In-game assistance-beta < In-game assistance-final	$p < 0.001$	139.994***	$\sim$ 5.188e−5
vrise-alpha < vrise-beta	$p = 0.111$	0.988	$\sim$ 0.001
VRISE-alpha <VRISE-final	$p < 0.001$	1912.328***	$\sim$ 3.643e-24
VRISE-beta < VRISE-final	$p < 0.001$	1277.335***	$\sim$ 7.819e-21

**BF$_{10}$ > 10; **BF$_{10}$ >30; ***BF$_{10}$ > 100; Alpha, Alpha version of VR-EAL; Beta, Beta version of VR-EAL; Final, Final Version of VR-EAL.*

and their probabilities against the corresponding null hypotheses are displayed in **Table 5**.

Moreover, the final version was substantially better than the alpha version in terms of every sub-score and total score of the VRNQ. The beta version was better than the alpha version in terms of the VRNQ total score as well as the user experience and game mechanics sub-scores. However, there was not a significant difference between the VRNQ in terms of the VRISE or in-game assistance sub-scores. Moreover, the final version was also significantly improved compared to the beta version in terms of the VRNQ total score and all sub-scores. Though, the difference between them was smaller in the game mechanics and user experience sub-scores (see **Table 5**). Importantly, in the final version of the VR-EAL, all users ($N = 25$) experienced mild (i.e., five in VRNQ) to no VRISE (i.e., seven in VRNQ), while the vast majority ($N = 22$) experienced very mild (i.e., six in VRNQ) to no VRISE (see **Figure 8**).

DISCUSSION

The VR-EAL Versions

The present study attempted to develop a cost-effective VR research/clinical software (i.e., VR-EAL) of a high enough quality for implementation in cognitive studies and that does not induce VRISE. The development included three versions of VR-EAL (i.e., alpha, beta, and final) until the attainment of these desired outcomes. The alpha version of VR-EAL revealed several limitations. It had low frames per second (fps), which increased the frequency and the intensity of VRISE. Also, the alpha version did not include haptics during the interactions, and the in-game assistance props were low in number. Lastly, the shaders of the 3D models were not converted to VR shaders (i.e., the function

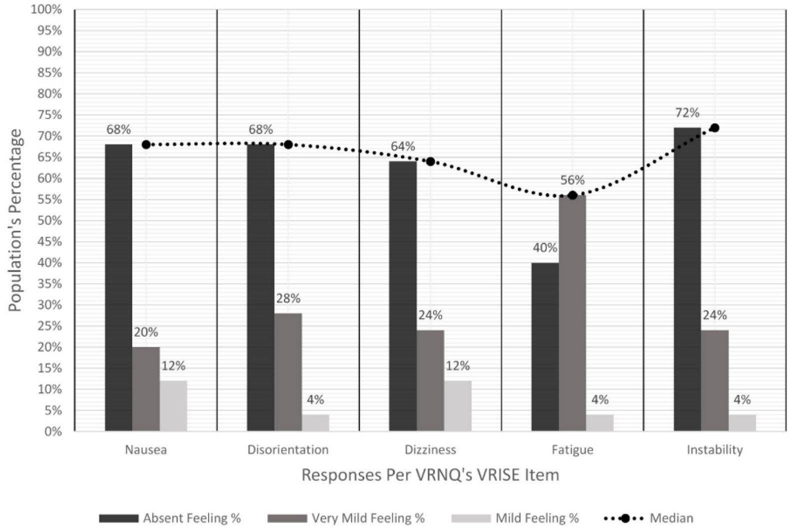


FIGURE 8 | VRISE in the final version of VR-EAL.

of the Lab renderer) and numerous game objects were defined as non-static. As a result, the quality of the graphics was below average, and the fps were substantially below 90 (i.e., 70–80) which is the lowest threshold for VR software targeting high-end HMDs such as HTC Vive and Oculus Rift. However, the feedback also confirmed that several game mechanics and approaches (e.g., tutorials) were in the right direction, which was encouraging for further VR-EAL development.

The principal improvements in the beta version of VR-EAL were pertinent to the alpha version's shortcomings. The shaders for all the game objects were converted to VR shaders, and several game objects, with which the user does not interact, were defined as static. The fps for the beta version were above 90, though, there were various points where the fps dropped for a couple of seconds. Although these fps drops were brief, their existence negatively affected the users who reported moderate to intense VRISE. Nonetheless, the beta version provided haptic and visual (i.e., highlighters) feedback to the users during the interactions, which further improved the quality of the game mechanics. In addition, the number of in-game aids was dramatically increased (e.g., more signs, labels, and directional arrows) and the duration of the tutorials was substantially prolonged (i.e., the inclusion of more explicit descriptions), which improved the quality of the users' experience. However, while the beta version was an improvement, it still failed to meet the parsimonious cut-offs of the VRNQ.

In the final version of VR-EAL, further improvements were conducted. The programming scripts of VR-EAL were re-assessed and correspondingly refined. Various chunks of code were expressed more compactly. For example, part of the code which had several Boolean values and/or float numbers were replaced by events and delegates (i.e., the features of object-oriented programming languages like C# that have substantially lower costs toward the performance of the software). Furthermore, the lightmapping of the 3D environments of scenes was upgraded by calculating high-resolution lightmaps instead of the medium resolution used in previous versions of VR-EAL. Redundant shadows were also deactivated to improve the performance of VR-EAL without degrading the quality of the graphics.

Moreover, major parts of the 3D environments were baked together (i.e., merged) through the implementation of MeshBaker's predominant functions to significantly reduce the draw-calls of VR-EAL. Interestingly, the result was a stable number of fps during gameplay. Specifically, the final version of VR-EAL has 120–140 fps during gameplay. Lastly, there was an improvement and enrichment of in-game assistance. In the tutorials, video screens and videos were added, which show the user how to use the controllers and perform each task. This visual and procedural demonstration allowed users to learn the respective controls and task trials faster and more effectively. This audio-visual demonstration using videos is feasible in VR since it can integrate the benefits of all mediums (e.g., video, audio, audio-visual). Furthermore, in the storyline scenes, where the user performs the actual tasks, several visual aids were added to provide additional guidance and alleviate confusion (see **Figures 3, 6**).

Our results demonstrated that the VRNQ total and sub-scores exceeded the parsimonious cut-offs of the VRNQ for the final VR-EAL version. The improvements pertinent to graphics substantially increased the quality of the user experience, while they almost eradicated VRISE (see **Figure 8**). This substantial decrease of VRISE also highlights the importance of fps in VR. A developer should use the Unity profiler to check whether the VR software has a steady number of fps during gameplay, which the HMD requires. Also, the final version of VR-EAL appeared to have better in-game assistance and game mechanics. However, there was not any upgrade pertinent to the game mechanics. The increase in the evaluation of the game mechanics probably resulted due to the addition and improvement of in-game aids in both tutorial and storyline scenes. This finding also supports that in-game assistance has a paramount role in VR software. This is especially the case when the software is developed for clinical or research purposes, where the users could be either gamers or non-gamers. The quality of the tutorials and in-game aids should be cautiously designed to ensure the usability of the VR research software. To sum up, the final version of VR-EAL seems to deliver a pleasant testing experience and without the presence of significant VRISE.

VR Software Development in Cognitive Sciences

The current study demonstrated the procedure for the development of immersive VR research/clinical software (i.e., VR-EAL) with strong placement and plausibility illusions, which are necessary for collecting genuine responses (i.e., ecological valid) from users (Slater, 2009; Slater et al., 2010; Maister et al., 2015; Pan and Hamilton, 2018). The implementation of good quality 3D models (e.g., objects, buildings, and artificial humans) in conjunction with optimization tools (e.g., Lab Renderer and MeshBaker) facilitated an analogous placement illusion. Also, VR-EAL incorporates naturalistic and ergonomic interactions with the VE facilitated by the VR hardware (e.g., HTC Vive and 6 DoF controllers), SDKs (e.g., SteamVR and VRTK), and Unity assets pertinent to spatialized audio (e.g., Steam Audio) and artificial characters' animations (e.g., Salsa3D). These naturalistic and ergonomic interactions with the VE are capable of inducing a robust plausibility illusion.

Furthermore, a predominant concern for the implementation of VR in cognitive sciences is the presence of VRISE (Bohil et al., 2011; de França and Soares, 2017; Palmisano et al., 2017), which may compromise health and safety standards (Parsons et al., 2018), as well as the reliability of cognitive (Nalivaiko et al., 2015), physiological (Nalivaiko et al., 2015), and neuroimaging data (Arafat et al., 2018; Gavgani et al., 2018). Equally, the high cost of VR software development may additionally deter the adoption of VR as a research tool in cognitive sciences (Slater, 2018). However, the development of VR-EAL provides evidence that the obstacles above can be surpassed to implement VR software in cognitive sciences effectively.

The users of the final version of VR-EAL reported mild to no VRISE, with the average value in the VRISE sub-score being very mild to no VRISE. Importantly, these reports were offered by the users after spending around 60 min uninterrupted in VR. Typically, VRISE are intensified in longer VR sessions (Sharples et al., 2008). However, the utilization of the parsimonious cut-offs from the VRNQ guaranteed the significant alleviation of VRISE, which was also supported by the users' reports. Notably, the results of this study are in line with our previous work (Kourtesis et al., 2019b), where the gaming experience (i.e., gamer or non-gamer) did not affect the responses on the VRNQ. Also, the results support that the gaming experience does not affect the presence or intensity of VRISE in software of adequate quality. Therefore, VR software with technical features similar to VR-EAL would be suitable for implementation in cognitive sciences.

Cognitive scientists already implement computational approaches to investigate cognitive functions at the neuronal and cellular level (Sejnowski et al., 1988; Farrell and Lewandowsky, 2010; Kriegeskorte and Douglas, 2018), develop computerized neuropsychological tasks compatible with neuroimaging techniques (Peirce, 2007, 2009; Mathôt et al., 2012), as well as conducting flexible statistical analyses and creating high-quality graphics and simulations (Culpepper and Aguinis, 2011; Revelle, 2011; Stevens, 2017). The development of VR-EAL was achieved by using C# and Unity packages (i.e., SteamVR SDK, VRTK, Lab renderer, MeshBaker, Salsa3D, RandomEyes3D, 3D models, 3D environments, and 3D characters) on the Unity game engine, which is a user-friendly IDE equivalent to OpenSesame, PsychoPy, and MATLAB.

The majority of these Unity packages are cost-free, while the remainder are relatively low-cost, and could be used in future VR software development. Also, the acquisition of VR development skills by cognitive scientists with a background in either psychology or computers science can be realized in a moderately short period. Although, collaboration with a psychologist who has the required knowledge and clinical experience is crucial for a computer scientist with VR skills. Likewise, psychologists should either collaborate with a computer scientist with VR expertise or acquire VR development skills themselves. For the acquisition of VR skills by a computer scientist or a psychologist, there are online and on-campus interdisciplinary modules (e.g., Unity tutorials and documentation, game development courses, programming workshops, and specializations in VR) which further support the feasibility of acquiring the necessary skills. However, training cognitive scientists in VR software development should be prioritized for institutions which aspire to implement VR technologies in their studies. To summarize, this study demonstrated that the development of usable VR research software by a cognitive scientist is viable.

Limitations and Future Studies

This study, however, has some limitations. The implementation of novel technologies may result in more positive responses toward them (Wells et al., 2010). A future replication of the current results would elucidate this issue. Also, the study did not provide validation of VR-EAL as a neuropsychological tool. Future work will consider validating the VR-EAL against traditional paper-and-pencil and computerized tests of prospective memory, executive function, episodic memory, and attentional processes. A future validation study should also include a larger and more diverse population than the sample in this study. Regarding the quality of VR-EAL, it is not able to induce a strong embodiment illusion. The future version of the VR-EAL should include a VR avatar that corresponds to the user's movements and actions. Also, the integration of better 3D models, environments, and characters may be beneficial, which will additionally improve the quality of placement illusion and the user's experience. Finally, since VR-EAL is ultimately intended for implementation in cognitive neuroscience and neuropsychology, the future version of VR-EAL should include compatibility with eye-tracking measurements and neuroimaging techniques (e.g., event-related potentials measured by electroencephalography).

Conclusion

This study provided guidelines for the development of immersive VR research software that can be implemented in cognitive sciences to improve the ecological validity of the cognitive tasks and automate the administration and scoring of the neuropsychological assessment. The results substantially support the feasibility of the development of low-cost and effective immersive VR software without the presence of VRISE during a 60 min VR session by cognitive scientists who have skills in VR software development. Technologically competent cognitive scientists are able to develop cost-effective immersive VR research software that guarantees the safety of the users and the reliability of the collected data (i.e., neuropsychological, physiological, and neuroimaging data).

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Philosophy, Psychology and Language Sciences Research Ethics Committee of the University of Edinburgh. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PK was the developer of VR-EAL. VR-EAL can be used by a third party by contacting the PK. PK had the initial idea and contributed to every aspect of this study. DK, SC, LD, and SM contributed to the methodological aspects and the discussion of the results.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2019.00012/full#supplementary-material

Supplementary Material III | A brief preview of VR-EAL.

REFERENCES

Anderson, F. T., McDaniel, M. A., and Einstein, G. O. (2017). Remembering to remember: an examination of the cognitive processes underlying prospective memory. *Learn. Mem.* 2, 451–463. doi: 10.1016/B978-0-12-809324-5.21049-3

Arafat, I. M., Ferdous, S. M. S., and Quarles, J. (2018). "Cybersickness-provoking virtual reality alters brain signals of persons with multiple sclerosis," In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Reutlingen: IEEE), 1–120. doi: 10.1109/VR.2018.8446194

Bohil, C. J., Alicea, B., and Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* 12:752. doi: 10.1038/nrn3122

Borrego, A., Latorre, J., Alcañiz, M., and Llorens, R. (2018). Comparison of oculus rift and HTC vive: feasibility for virtual reality-based exploration, navigation, exergaming, and rehabilitation. *Games Health J.* 7, 151–156. doi: 10.1089/g4h.2017.0114

Bozgeyikli, E., Raij, A., Katkoori, S., and Dubey, R. (2016). "Point and teleport locomotion technique for virtual reality," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (ACM), 205–216. doi: 10.1145/2967934.2968105

Brade, J., Dudczig, M., and Klimant, P. (2018). "Using virtual prototyping technologies to evaluate human-machine-interaction concepts," in *AW&I Conference*, Vol. 3 (Chemnitz).

Brennesholtz, M. S. (2018). "3-1: invited paper: vr standards and guidelines," in *SID Symposium Digest of Technical Papers,* Vol. 49 (Pleasantville, NY), 1–4. doi: 10.1002/sdtp.12476

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., and Wilson, B. A. (1998). The ecological validity of tests of executive function. *J. Int. Neuropsychol. Soc.* 4, 547–558. doi: 10.1017/S1355617798466037

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

Caputo, F., Greco, A., Egidio, D. A., Notaro, I., and Spada, S. (2017). "A preventive ergonomic approach based on virtual and immersive reality," in *International Conference on Applied Human Factors and Ergonomics* (Cham: Springer), 3–15. doi: 10.1007/978-3-319-60582-1_1

Chandra, K. (2012). A comparison of objective-C and C#: implementing Cbox class. *J. Comput. Sci. Coll.* 27, 178–186.

Chaytor, N., and Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. *Neuropsychol. Rev.* 13, 181–197. doi: 10.1023/B:NERV.0000009483.91468.fb

Cox, D. R., and Donnelly, C. A. (2011). *Principles of Applied Statistics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139005036

Culpepper, S. A., and Aguinis, H. (2011). R is for revolution: a cutting-edge, free, open source statistical package. *Org. Res. Methods* 14, 735–740. doi: 10.1177/1094428109355485

de França, A. C. P., and Soares, M. M. (2017). "Review of virtual reality technology: an ergonomic approach and current challenges," in *International Conference on Applied Human Factors and Ergonomics* (Cham: Springer), 52–61. doi: 10.1007/978-3-319-60582-1_6

Dickson, P. E., Block, J. E., Echevarria, G. N., and Keenan, K. C. (2017). "An experience-based comparison of unity and unreal for a stand-alone 3D game development course," in *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education* (ACM), 70–75. doi: 10.1145/3059009.3059013

Elkind, J. S., Rubin, E., Rosenthal, S., Skoff, B., and Prather, P. (2001). A simulated reality scenario compared with the computerized Wisconsin card sorting test: an analysis of preliminary results. *CyberPsychol. Behav.* 4, 489–496. doi: 10.1089/109493101750527042

Farrell, S., and Lewandowsky, S. (2010). Computational models as aids to better reasoning in psychology. *Curr. Dir. Psychol. Sci.* 19, 329–335. doi: 10.1177/0963721410386677

Figueiredo, L., Rodrigues, E., Teixeira, J., and Techrieb, V. (2018). A comparative evaluation of direct hand and wand interactions on consumer devices. *Comput. Graphics* 77, 108–121. doi: 10.1016/j.cag.2018.10.006

Frommel, J., Sonntag, S., and Weber, M. (2017). "Effects of controller-based locomotion on player experience in a virtual reality exploration game," in *Proceedings of the 12th International Conference on the Foundations of Digital Games* (ACM), 30. doi: 10.1145/3102071.3102082

Gavgani, A. M., Wong, R. H. X., Howe, P. R. C., Hodgson, D. M., Walker, F. R., and Nalivaiko, E. (2018). Cybersickness-related changes in brain hemodynamics: a pilot study comparing transcranial Doppler and near-infrared spectroscopy assessments during a virtual ride on a roller coaster. *Physiol. Behav.* 191, 56–64. doi: 10.1016/j.physbeh.2018.04.007

Gehrke, L., Iversen, J. R., Makeig, S., and Gramann, K. (2018). "The invisible maze task (IMT): interactive exploration of sparse virtual environments to investigate action-driven formation of spatial representations," in *German Conference on Spatial Cognition* (Cham: Springer), 293–310. doi: 10.1007/978-3-319-96385-3_20

Gonzalez-Franco, M., and Lanier, J. (2017). Model of illusions and virtual reality. *Front. Psychol.* 8:1125. doi: 10.3389/fpsyg.2017.01125

Grewe, P., Lahr, D., Kohsik, A., Dyck, E., Markowitsch, H. J., Bien, C. G., et al. (2014). Real-life memory and spatial navigation in patients with focal epilepsy: ecological validity of a virtual reality supermarket task. *Epilepsy Behav.* 31, 57–66. doi: 10.1016/j.yebeh.2013.11.014

Gromala, D., Tong, X., Shaw, C., Amin, A., Ulas, S., and Ramsay, G. (2016). Mobius floe: an immersive virtual reality game for pain distraction. *Electr. Imaging* 2016, 1–5. doi: 10.2352/ISSN.2470-1173.2016.4.ERVR-413

Haines, S., Shelton, J., Henry, J., Terrett, G., Vorwerk, T., and Rendell, P. (2019). *Prospective Memory and Cognitive Aging.* Oxford Research Encyclopedia of Psychology. Available online at: https://oxfordre.com/psychology/view/10.1093/acrefore/9780190236557.001.0001/acrefore-9780190236557-e-381 (accessed December 7, 2019).

Held, L., and Ott, M. (2018). On p-values and Bayes factors. *Ann. Rev. Stat. Appl.* 5, 393–419. doi: 10.1146/annurev-statistics-031017-100307

Higginson, C. I., Arnett, P. A., and Voss, W. D. (2000). The ecological validity of clinical tests of memory and attention in multiple sclerosis. *Arch. Clin. Neuropsychol.* 15, 185–204. doi: 10.1093/arclin/15.3.185

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

JASP Team (2017). *JASP (Version 0.8.1.2)* [Computer software]. JASP Team

Jerald, J. (2015). *The VR Book: Human-Centered Design for Virtual Reality.* New York, NY: Morgan and Claypool. doi: 10.1145/2792790

Jerald, J., LaViola Jr, J. J., and Marks, R. (2017). "VR interactions," in *ACM SIGGRAPH 2017 Courses* (Los Angeles, CA: ACM), 19. doi: 10.1145/3084873.3084900

Kobayashi, M., Ueno, K., and Ise, S. (2015). The effects of spatialized sounds on the sense of presence in auditory virtual environments: a psychological and physiological study. *Presence* 24, 163–174. doi: 10.1162/PRES_a_00226

Korre, D. (2019). *Usability evaluation of spoken humanoid embodied conversational agents in mobile serious games* (PhD thesis). University of Edinburgh, Edinburgh, United Kingdom.

Kourtesis, P., Collina, S., Doumas, L. A. A., and MacPherson, S. E. (2019a). Technological competence is a precondition for effective implementation of virtual reality head mounted displays in human neuroscience: a technological review and meta-analysis. *Front. Hum. Neurosci.* 13:342. doi: 10.3389/fnhum.2019.00342

Kourtesis, P., Collina, S., Doumas, L. A. A., and MacPherson, S. E. (2019b). Validation of the virtual reality neuroscience questionnaire: maximum duration of immersive virtual reality sessions without the presence of pertinent adverse symptomatology. *Front. Hum. Neurosci.* 13:417. doi: 10.3389/fnhum.2019.00417

Kriegeskorte, N., and Douglas, P. K. (2018). Cognitive computational neuroscience. *Nat. Neurosci.* 21, 1148–1160. doi: 10.1038/s41593-018-0210-5

LaViola, J. J. Jr., Kruijff, E., McMahan, R. P., Bowman, D., and Poupyrev, I. P. (2017). *3D User Interfaces: Theory and Practice.* Boston, MA: Addison-Wesley Professional.

Lavoué, G., and Mantiuk, R. (2015). "Quality assessment in computer graphics," in *Visual Signal Quality Assessment* (Cham: Springer), 243–286. doi: 10.1007/978-3-319-10368-6_9

Maister, L., Slater, M., Sanchez-Vives, M. V., and Tsakiris, M. (2015). Changing bodies changes minds: owning another body affects social cognition. *Trends Cogn. Sci.* 19, 6–12. doi: 10.1016/j.tics.2014.11.001

Makeig, S., Gramann, K., Jung, T. P., Sejnowski, T. J., and Poizner, H. (2009). Linking brain, mind and behavior. *Int. J. Psychophysiol.* 73, 95–100. doi: 10.1016/j.ijpsycho.2008.11.008

Marsman, M., and Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *Euro. J. Dev. Psychol.* 14, 545–555. doi: 10.1080/17405629.2016.1259614

Mathôt, S., Schreij, D., and Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behav. Res. Methods* 44, 314–324. doi: 10.3758/s13428-011-0168-7

McDaniel, M. A., and Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: a multiprocess framework. *Appl. Cogn. Psychol.* 14, S127–S144. doi: 10.1002/acp.775

McDaniel, M. A., and Einstein, G. O. (2007). *Prospective Memory: An Overview and Synthesis of an Emerging Field.* Thousand Oaks, CA: Sage Publications.

Mlinac, M. E., and Feng, M. C. (2016). Assessment of activities of daily living, self-care, and independence. *Arch. Clin. Neuropsychol.* 31, 506–516. doi: 10.1093/arclin/acw049

Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robot. Automat. Mag.* 19, 98–100. doi: 10.1109/MRA.2012.2192811

Nalivaiko, E., Davis, S. L., Blackmore, K. L., Vakulin, A., and Nesbitt, K. V. (2015). Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiol. Behav.* 151, 583–590. doi: 10.1016/j.physbeh.2015.08.043

Palmisano, S., Mursic, R., and Kim, J. (2017). Vection and cybersickness generated by head-and-display motion in the Oculus Rift. *Displays* 46, 1–8. doi: 10.1016/j.displa.2016.11.001

Pan, X., and Hamilton, A. F. D. C. (2018). Why and how to use virtual reality to study human social interaction: the challenges of exploring a new research landscape. *Br. J. Psychol.* 109, 395–417. doi: 10.1111/bjop.12290

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front. Hum. Neurosci.* 9:660. doi: 10.3389/fnhum.2015.00660

Parsons, T. D., McMahan, T., and Kane, R. (2018). Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms. *Clin. Neuropsychol.* 32, 16–41. doi: 10.1080/13854046.2017.1337932

Peirce, J. W. (2007). PsychoPy–psychophysics software in Python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinf.* 2:10. doi: 10.3389/neuro.11.010.2008

Phillips, L. H., Henry, J. D., and Martin, M. (2008). "Adult aging and prospective memory: the importance of ecological validity," in *Prospective Memory: Cognitive, Neuroscience, Developmental, and Applied Perspectives*, eds M. Kliegel, M. A. McDaniel, and G. O. Einstein (Taylor & Francis Group, Lawrence Erlbaum Associates), 161–185.

Porcino, T. M., Clua, E., Trevisan, D., Vasconcelos, C. N., and Valente, L. (2017). "Minimizing cyber sickness in head mounted display systems: design guidelines and applications," in *Serious Games and Applications for Health (SeGAH), 2017 IEEE 5th International Conference on* (IEEE), 1–6. doi: 10.1109/SeGAH.2017.7939283

Rand, D., Basha-Abu Rukan, S., Weiss, P. L., and Katz, N. (2009). Validation of the Virtual MET as an assessment tool for executive functions. *Neuropsychol. Rehab.* 19, 583–602. doi: 10.1080/09602010802469074

Revelle, W. (2011). *An Overview of the Psych Package*. Evanston, IL: Department of Psychology, Northwestern University.

Rizzo, A. A., Schultheis, M., Kerns, K. A., and Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychol. Rehab.* 14, 207–239. doi: 10.1080/09602010343000183

Robertson, I. H., Ward, T., Ridgeway, V., and Nimmo-Smith, I. (1994). *The Test of Everyday Attention (TEA)*. Bury St. Edmunds: Thames Valley Test Company.

Rosenberg, L. (2015). The associations between executive functions' capacities, performance process skills, and dimensions of participation in activities of daily life among children of elementary school age. *Appl. Neuropsychol.* 4, 148–156. doi: 10.1080/21622965.2013.821652

Rumiński, D. (2015). An experimental study of spatial sound usefulness in searching and navigating through AR environments. *Virt. Real.* 19, 223–233. doi: 10.1007/s10055-015-0274-4

Sejnowski, T. J., Koch, C., and Churchland, P. S. (1988). Computational neuroscience. *Science* 241, 1299–1306. doi: 10.1126/science.3045969

Seyama, J. I., and Nagayama, R. S. (2007). The uncanny valley: effect of realism on the impression of artificial human faces. *Presence* 16, 337–351. doi: 10.1162/pres.16.4.337

Sharples, S., Cobb, S., Moody, A., and Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): comparison of head mounted display (HMD), desktop and projection display systems. *Displays* 29, 58–69. doi: 10.1016/j.displa.2007.09.005

Sherman, W. R., and Craig, A. B. (2018). *Understanding Virtual Reality: Interface, Application, and Design*. San Francisco, CA: Morgan Kaufmann. doi: 10.1016/B978-0-12-800965-9.00010-6

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. B* 364, 3549–3557. doi: 10.1098/rstb.2009.0138

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *Br. J. Psychol.* 109:431. doi: 10.1111/bjop.12305

Slater, M., Spanlang, B., and Corominas, D. (2010). Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Trans. Graphics* 29:92. doi: 10.1145/1778765.1778829

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 347. doi: 10.1037/0278-7393.29.3.347

Smith, R. E., Hunt, R. R., McVay, J. C., and McConnell, M. D. (2007). The cost of event-based prospective memory: Salient target events. *J. Exp. Psychol.* 33:734. doi: 10.1037/0278-7393.33.4.734

Sportillo, D., Paljic, A., Boukhris, M., Fuchs, P., Ojeda, L., and Roussarie, V. (2017). "An immersive Virtual Reality system for semi-autonomous driving simulation: a comparison between realistic and 6-DoF controller-based interaction," in *Proceedings of the 9th International Conference on Computer and Automation Engineering* (Sydney: ACM), 6–10. doi: 10.1145/3057039.3057079

Stevens, J. R. (2017). Replicability and reproducibility in comparative psychology. *Front. Psychol.* 8:862. doi: 10.3389/fpsyg.2017.00862

Toschi, N., Kim, J., Sclocco, R., Duggento, A., Barbieri, R., Kuo, B., et al. (2017). Motion sickness increases functional connectivity between visual motion and nausea-associated brain regions. *Auton. Neurosci.* 202, 108–113. doi: 10.1016/j.autneu.2016.10.003

Viirre, E., Price, B. J., and Chase, B. (2014). "Direct effects of virtual environments on users," in *Handbook of Virtual Environments* (CRC Press), 527–535. doi: 10.1201/b17360-28

Vorländer, M., and Shinn-Cunningham, B. (2014). "Virtual auditory displays," in *Handbook of Virtual Environments* (CRC Press), 107–134. doi: 10.1201/b17360-7

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., et al. (2018a). Bayesian inference for psychology. Part II: example applications with JASP. *Psychon. Bull. Rev.* 25, 58–76. doi: 10.3758/s13423-017-1323-7

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., et al. (2018b). Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. *Psychon. Bull. Rev.* 25, 35–57. doi: 10.3758/s13423-017-1343-3

Wells, J. D., Campbell, D. E., Valacich, J. S., and Featherman, M. (2010). The effect of perceived novelty on the adoption of information technology innovations: a risk/reward perspective. *Dec. Sci.* 41, 813–843 doi: 10.1111/j.1540-5915.2010.00292.x

Wilson, B. A., Cockburn, J., and Baddeley, A. (2008). *The Rivermead Behavioural Memory Test*. London: Pearson Assessment.

Wilson, B. A., Emslie, H., Foley, J., Shiel, A., Watson, P., Hawkins, K., et al. (2005). *The Cambridge Prospective Memory Test: CAMPROMPT*. London: Harcourt Assessment.

Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., and Burgess, P. (1998). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychol. Rehab.* 8, 213–228. doi: 10.1080/713755570

Zaidi, S. F. M., Duthie, C., Carr, E., and Maksoud, S. H. A. E. (2018). "Conceptual framework for the usability evaluation of gamified virtual reality environment for non-gamers," in *Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry* (Tokyo: ACM), 13. doi: 10.1145/3284398.3284431

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

This chapter provided guidelines for the development of immersive VR software for implantation in cognitive neuroscience and neuropsychology. Also, this chapter demonstrated how the VRNQ may be used for the assessment of the VR software features and the intensity of VRISE. Importantly, Chapter 4 showed that the VR-EAL provides an adequate level of immersion, a pleasant testing experience, and it does not induce intense VRISE. However, the psychometric properties of the VR-EAL were not explored in this chapter. Hence, the validity and reliability of VR-EAL should be investigated.

Chapter 5: Validation of the Virtual Reality Everyday Assessment Lab (VR-EAL): An Immersive Virtual Reality Neuropsychological Battery with Enhanced Ecological Validity

---

Chapter 4 showed that VR-EAL is an innocuous immersive VR software regarding VRISE intensity, as well as that the VR-EAL offers a deep immersion and a pleasant testing experience. However, the psychometric properties were not examined. In Chapter 5, the construct, content, and ecological validity will be investigated, as well as its internal reliability. Furthermore, a comparison in terms of pleasantness, administration time, and similarity to the everyday life tasks between the VR-EAL and an extensive paper-and-pencil neuropsychological battery will be conducted in order the advantages of the VR-EAL and using immersive VR technologies to be scrutinised.

This chapter has been accepted for publication by the Journal of the International Neuropsychological Society: Kourtesis, P., Collina, S., Doumas, L. A., & MacPherson, S. E. (2020). Validation of the Virtual Reality Everyday Assessment Lab (VR-EAL): An immersive virtual reality neuropsychological battery with enhanced ecological validity. *Journal of the International Neuropsychological Society*, in press

---

Validation of the Virtual Reality Everyday Assessment Lab (VR-EAL): An immersive virtual reality neuropsychological battery with enhanced ecological validity

Panagiotis Kourtesis[a,b,c,d]*, Simona Collina[c,d], Leonidas A.A. Doumas[b], and Sarah E. MacPherson[a,b].

[a] *Human Cognitive Neuroscience, Department of Psychology, University of Edinburgh, Edinburgh, UK;*
[b] *Department of Psychology, University of Edinburgh, Edinburgh, UK;*

[c] *Lab of Experimental Psychology, Suor Orsola Benincasa University of Naples, Naples, Italy;*

[d] *Interdepartmental Centre for Planning and Research "Scienza Nuova", Suor Orsola Benincasa University of Naples, Naples, Italy*

* Panagiotis Kourtesis, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, Scotland, United Kingdom

Email: pkourtes@exseed.ed.ac.uk

Abstract

Objective: The assessment of cognitive functions such as prospective memory, episodic memory, attention, and executive functions benefits from an ecologically valid approach to better understand how performance outcomes generalize to everyday life. Immersive virtual reality (VR) is considered capable of simulating real-life situations to enhance ecological validity. The present study attempted to validate the Virtual Reality Everyday Assessment Lab (VR-EAL), an immersive VR neuropsychological battery, against an extensive paper-and-pencil neuropsychological battery.

Methods: Forty-one participants (21 females) were recruited: 18 gamers and 23 non-gamers who attended both an immersive VR and a paper-and-pencil testing session. Bayesian Pearson correlation analyses were conducted to assess construct and convergent validity of the VR-EAL. Bayesian t-tests were performed to compare VR and paper-and-pencil testing in terms of administration time, similarity to real life tasks (i.e., ecological validity), and pleasantness.

Results: VR-EAL scores were significantly correlated with their equivalent scores on the paper-and-pencil tests. The participants' reports indicated that the VR-EAL tasks were significantly more ecologically valid and pleasant than the paper-and-pencil neuropsychological battery. The VR-EAL battery also had a shorter administration time.

Conclusion: The VR-EAL appears as an effective neuropsychological tool for the assessment of everyday cognitive functions, which has enhanced ecological validity, a highly pleasant testing experience, and does not induce cybersickness.

Keywords: Prospective Memory, Episodic Memory, Attention, Executive Function, Everyday Functioning, Virtual Reality.

Introduction

The ability to perform activities in everyday life is dependent upon cognitive abilities such as attention, episodic memory, executive abilities and prospective memory (Mlinac & Feng, 2016). The neuropsychological assessment of these cognitive abilities benefits from an ecologically valid approach to better understand the quality of an individual's everyday functioning (Chaytor & Schmitter-Edgecombe, 2003). Ecological validity increases the probability that an individual's cognitive performance will replicate how they will respond in real-life situations (Bailey, Henry, Rendell, Phillips, & Kliegel, 2010; Burgess *et al.*, 2006; Chaytor & Schmitter-Edgecombe, 2003).

Verisimilitude and veridicality are the two predominant approaches for achieving the ecological validity of neuropsychological tests (Franzen & Wilhelm, 1996; Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). Verisimilitude refers to the level of resemblance to the complexity and cognitive demands of everyday tasks by the neuropsychological tests (Franzen & Wilhelm, 1996; Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). Veridicality refers to the strength of the relationship between the outcomes of neuropsychological tests and everyday functioning measures (e.g., questionnaires pertinent to everyday functioning and independence; Franzen & Wilhelm, 1996; Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). While both verisimilitude and veridicality approaches have their merits, the literature suggests that the verisimilitude approach may be better predictors of real-world memory and attention (Higginson, Arnett, & Voss, 2000), executive functioning (e.g., multi-tasking, planning and mental flexibility; Burgess, Alderman, Evans, Emslie, &

Wilson, 1998) and prospective memory abilities (e.g., remembering to initiate a planned action in the future; Haines *et al.*, 2019; Phillips, Henry, & Martin, 2012) than the veridicality approach. (Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006).

Several laboratory-based test batteries that simulate real life tasks exist in the neuropsychological literature including those assessing attention (e.g., Test of Everyday Attention, TEA; Robertson, Ward, Ridgeway, & Nimmo-Smith, 1996), memory (e.g., Rivermead Behavioral Memory Test–III, RBMT–III; Wilson, Cockburn, & Baddeley, 2008), executive abilities (e.g., Behavioral Assessment of Dysexecutive Syndrome, BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1997) and prospective memory (e.g., Cambridge Prospective Memory Test, CAMPROMPT; Wilson, 2005). Yet, such neuropsychological test batteries tend to incorporate simple, static stimuli within a highly controlled environment and do not fully resemble the complexity of real-life situations (Parsons, 2015; Rand, Rukan, Weiss, & Katz, 2009). Attempts to provide better assessments of everyday abilities have involved assessments in real-life settings such as performing errands in a shopping center or a pedestrianized street (e.g., Garden, Phillips, & MacPherson, 2001; Shallice & Burgess, 1991). However, these cannot be standardized for use in other clinics or laboratories, they may not be feasible for some individuals in challenging populations (e.g., psychiatric patients, stroke patients with paresis or paralysis), they are time-consuming and expensive, they require participant transport and consent from local businesses and they lack experimental control over the external situation (e.g., Elkind, Rubin, Rosenthal, Skoff, & Prather, 2001; Logie, Trawley, & Law, 2011; Parsons, 2015; Rand *et al.*, 2009).

The use of technology such as video recordings of real-world locations and non-immersive virtual environments (Farrimond, Knight, & Titov, 2006; McGeorge *et al.*, 2001; Paraskevaides *et al.*, 2010) have also been considered to simulate real-life situations. Non-immersive virtual reality (VR) tests such as the Edinburgh Virtual Errands Test (EVET; Logie *et al.*, 2011), the Jansari Assessment of Executive Function (Jansari *et al.*, 2014), the Virtual Multiple Errands Test (VMET) within the Virtual Mall (VMall; Rand *et al.*, 2009) and the Virtual Reality Shopping Task (Canty *et al.*, 2014) attempt to simulate real-life tasks and are considered more cost-effective, require less administration time, have greater experimental control and can be easily be adapted for other clinical or research settings (Parsons, McMahan, & Kane, 2018; Werner & Korczyn, 2012; Zygouris & Tsolaki, 2015). Non-immersive VR tests can also offer automated scoring and standardized administration, enabling clinicians and researchers to administer these tests with only limited training. Finally, some non-immersive VR tests also offer shorter versions of the test that focus on the assessment of specific cognitive functions (Parsons *et al.*, 2018; Werner & Korczyn, 2012; Zygouris & Tsolaki, 2015).

However, the user interface and procedure of non-immersive VR tests can be challenging for individuals without gaming backgrounds (Parsons *et al.*, 2018; Zaidi, Duthie, Carr, & Maksoud, 2018), especially for older adults and clinical populations such as individuals with mild cognitive impairment or Alzheimer's disease (Werner & Korczyn, 2012; Zygouris & Tsolaki, 2015). Immersive VR tests, which share the same advantages as non-immersive ones, may overcome these challenges (Rizzo, Schultheis, Kerns, & Mateer, 2004; Bohil, Alicea, & Biocca, 2011; Parsons, 2015; Teo *et al.*, 2016). In addition, individuals without gaming experience have been found to perform better in immersive VR environments due to the first-person perspective and ergonomic/naturalistic interactions that are proximal to real-life actions (Zaidi *et al.*, 2018). Also, while VR tests have in the past resulted in VR-induced symptoms and effects (VRISE) such as nausea, dizziness, disorientation, fatigue, or instability (Bohil *et al.*, 2011; de Franca & Soares, 2017; Palmisano, Mursic, & Kim, 2017), which compromise neuropsychological (Mittelstaedt, Wacker, & Stelling, 2018; Nalivaiko, Davis, Blackmore, Vakulin, & Nesbitt, 2015; Nesbitt, Davis, Blackmore, & Nalivaiko, 2017) and neuroimaging data (Arafat, Ferdous, & Quarles, 2018; Gavgani *et al.*, 2018; Toschi *et al.*, 2017), certain contemporary VR head-mounted displays (HMDs) and VR software with naturalistic and ergonomic interactions and navigation within the virtual environment reduce or show no symptoms of VRISE (see Kourtesis, Collina, Doumas, & MacPherson, 2019a). Lastly, immersive VR has been found to provide deeper immersion in the virtual environment than non-immersive VR; deeper immersion has been found to induce substantially less adverse VRISE (Kourtesis, Collina, Doumas, & MacPherson, 2019b; Weech, Kenny, & Barnett-Cowan, 2019).

We recently developed the Virtual Reality Everyday Assessment Lab (VR-EAL) to create an immersive virtual environment that simulates everyday tasks proximal to real-life to assess prospective memory, episodic memory (immediate and delayed recognition), executive functions (i.e., multitasking and planning) and selective visual, visuospatial and auditory attention (Kourtesis, Korre, Collina, Doumas, & MacPherson, 2020). In the VR-EAL, individuals are exposed to alternating tutorials (practice trials) and storyline tasks (assessments) to allow them to become familiarized with both the immersive VR technology and the specific controls and procedures of each VR-EAL task. Moreover, VR-EAL offers also a shorter version (i.e., scenario) where only episodic memory, executive function, selective visual attention, and selective visuospatial attention are assessed. Also, the examiner can opt to simply assess a specific cognitive function, where the examinee will go through the generic tutorial, the specific tutorial for this task, and the storyline task that assess the chosen cognitive function (e.g., selective visual attention).

VR-EAL endeavors to be the first immersive VR neuropsychological battery of everyday cognitive functions. Our previous work has shown that the VR-EAL does not induce VRISE (Kourtesis *et al.*, 2020). However, we have yet to demonstrate the validity of the VR-EAL as a neuropsychological tool. In the current study, the full version of the VR-EAL was administered to participants and compared with existing paper-and-pencil neuropsychological tests to assess the construct validity of the VR-EAL. We also aimed to replicate our previous findings that the VR-EAL does not induce VRISE, using the virtual reality neuroscience questionnaire (VRNQ; Kourtesis *et al.*, 2019b). Finally, comparisons between the VR-EAL and neuropsychological paper-and-pencil tests were conducted in terms of verisimilitude (i.e., ecological validity), pleasantness, and administration time.

Methods

Participants

Participants were recruited via social media and the internal mailing list of the University of Edinburgh. Forty-one participants (21 females) aged between 18 and 45 years (M = 29.15, SD = 5.80) were recruited: 18 considered themselves to be gamers (7 females) and 23 (14 females) considered themselves to be non-gamers. The mean education of the group was 13.80 years (SD = 2.36, range = 10-16). The study was approved by the Philosophy, Psychology and Language Sciences Research Ethics Committee of the University of Edinburgh. Written informed consent was obtained from each participant. All participants received verbal and written instructions regarding the procedures, possible adverse effects of immersive VR (e.g., VRISE), utilization of the data, and general aims of the study.

Materials

Hardware. An HTC Vive HMD with two lighthouse stations for motion tracking and two HTC Vive wands with six degrees of freedom (6DoF) for navigation and interactions within the virtual environment were implemented in accordance with our previously published technological recommendations for immersive VR research (Kourtesis *et al.*, 2019a). The spatialized (bi-aural) audio was facilitated by a pair of Senhai Kotion Each G9000 headphones. The size of the VR area was 5m^2, which facilitates an adequate space for immersion and naturalistic interaction within virtual environments (Borrego, Latorre, Alcañiz, & Llorens, 2018). The HMD was connected to a laptop with an Intel Core i7 7700HQ 2.80GHz processor, 16 GB RAM, a 4095MB NVIDIA GeForce GTX 1070 graphics card, a 931 GB TOSHIBA MQ01ABD100 (SATA) hard disk, and Realtek High Definition Audio.

VR-EAL. VR-EAL attempts to assess everyday cognitive functioning by assessing prospective memory, episodic memory (i.e., immediate and delayed recognition), executive functioning (i.e., planning, multitasking) and selective visual, visuospatial and auditory (bi-aural) attention within a realistic immersive VR scenario lasting around 60 minutes (Kourtesis *et al.*, 2020). See Table 1 and Figures 1 and 2 for a summary of the VR-EAL tasks assessing each cognitive ability. See Table 2 for the description of the VR-EAL tasks and Table 3 for the administration procedures and scoring of the VR-EAL tasks. For a full description of the VR-EAL's scenarios, tasks, and scoring, see Kourtesis *et al.* (2020). Also, a brief video recording of the VR-EAL may be accessed at this hyperlink: https://www.youtube.com/watch?v=IHEIvS37Xy8&t= .

Table 1. VR-EAL tasks and score ranges

Scene	Cognitive Function	Task	Score Ranges
3	Prospective memory	Write down the notes for the errands.	0 – 6
3	Immediate recognition	Recognising items on the shopping list.	0 – 20
3	Planning	Drawing the route to be taken.	0 – 19
6	Multitasking	Cooking task (preparing breakfast).	0 – 16
6	Prospective memory – event based	Take medication after breakfast.	0 – 6
8	Selective visuospatial attention	Collect items from the living room.	0 – 20
8	Prospective memory – event based	Take the chocolate pie out of the oven.	0 – 6
10	Prospective memory – time based	Call Rose at 10 am.	0 – 6
12	Selective visual attention	Find posters on both sides of the road.	0 – 16
14	Delayed recognition	Recognising items from the shopping list.	0 – 20
15	Prospective memory – time based	Collect the carrot cake from the bakery at 12 pm.	0 – 6
16	Prospective memory – event based	False prompt before going to the library.	-6 – 0
17	Prospective memory – event based	Return the red book to the library.	0 – 6
19	Selective auditory attention	Detect sounds from both sides of the road.	0 – 32
20	Prospective memory – time based	False prompt before going back home.	-6 – 0
21	Prospective memory – event based	Back home, give the extra pair of keys to Alex.	0 – 6
22	Prospective memory – time based	Take the medication at 1pm.	0 – 6

*The tasks are presented in the same order as they are performed within the scenario.

Table 2. VR-EAL tasks' description

Cognitive Function	**Description**
Episodic memory	Both immediate and delayed episodic memory are assessed. Firstly, the participant needs to memorize a shopping list which is presented audio-visually. Immediately after the presentation of the list, the participant is presented with 30 items and should visually recognize and select the 10 items from the shopping list (immediate recognition). Participants are then expected to choose the items from the list when they arrive at the supermarket approximately 20 minutes later (delayed recognition).
Executive Function: Planning	Planning ability is assessed by asking the participant to draw his or her route around the city (e.g., visiting the bakery, supermarket, library, and returning home) on a 3D interactive board.
Executive Function: Multitasking	Multitasking is examined using a cooking task, where the participant should prepare and serve his or her breakfast (e.g., sausages, omelet, and a cup of tea/coffee) and place a chocolate pie in the oven.
Prospective Memory	Comparable to the CAMPROMPT, VR-EAL considers both event-based and time-based prospective memory tasks. In the event-based tasks, the participant should remember to perform a prospective memory action when a particular event occurs (e.g., take medicines after breakfast). In the time-based tasks, the examinee should remember to perform a planned action at a specific time (e.g., call Rose at 12 pm).
Visuospatial Attention	Visuospatial attention is assessed by asking the participant to find and collect 6 specific items (i.e., a mobile phone, a £50 note, a library card, the flat keys, a red book, and car keys) in the living-room. A reminder of these items remains on the wall (i.e., the items are displayed as 3D objects with labels). However, there are also distractors (i.e., magazines, books, a remote control, a notebook, a pencil, a chessboard, and a bottle of wine) in the room.
Visual Attention	Visual attention is measured while the participant is seated as a passenger in a car next to a driver. The participant should identify all the targets (i.e., 16 posters of a radio station) on both sides of the road, while s/he needs to avoid any distractors (i.e., 8 posters that are a different shape and 8 posters with a different background color).
Auditory Attention	Auditory attention is also examined while the participant is seated as a passenger next to a driver. The participant should detect all the target sounds (i.e., 16 bell sounds) presented on both sides of the road, while avoiding the distractor sounds (i.e., 8 high pitch bells, and 8 dongs).

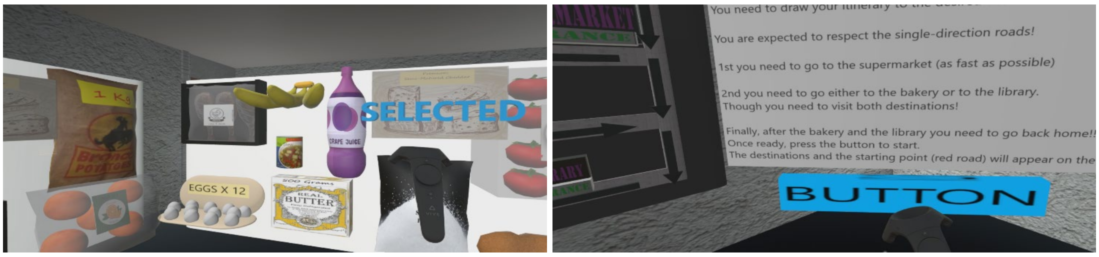
Table 3. VR-EAL task administration and scoring

Task	Scoring
Episodic memory	The user should choose the ten target items (i.e., create the shopping list) from an extensive array of items, which also contains five qualitative distractors (e.g., semi-skimmed milk versus skimmed milk), five quantitative distractors (e.g., 1 kg potatoes versus 2 kg potatoes), and ten false items (e.g., bread, bananas etc.). The user gains 2 points for each correctly chosen item, 1 point for choosing a qualitative or quantitative distractor, and 0 points for the false items. Scores range from 0 to 20.
Planning	The road system comprises 23 street units. When the user selects a unit, 1 point is awarded. The ideal route to visit all three destinations is 15 units; hence, any extra or missing units are subtracted from the total possible score of 15. Up to 4 more points are awarded for the time taken to complete the task. Scores range from 0 to 19.
Multitasking	Scoring relies on the animations from each game object (i.e., the omelet and the sausages). At the beginning of the animation, both items have a reddish (raw) color which gradually turns to either a yellowish (omelet) or brownish (sausages) color, and finally both turn to black (burnt). The score for each pan hence depends on the time that the user removes the pans from the stove and places them on the kitchen worktop. Equally, the score for boiling the kettle is measured in relation to the stage of the audio playback (e.g., the kettle whistles when the water is ready) that the kettle is placed on the kitchen worktop. Scores range from 0 to 16.
Prospective Memory	**Example:** At the end of a scene, the user should press a button to confirm that all the tasks in the scene are completed. If the user has already taken his/her medication (i.e., prospective memory task) before pressing the final button, then the scene ends, and the user receives 6 points. Otherwise, the first prompt appears (i.e., "You Have to Do Something Else"). If the user then follows the prompt and takes their medication, they receive 4 points. If the user presses the final button again, then the second prompt appears (i.e., "You Have to Do Something After Having your Breakfast"). If the user follows this prompt and takes their medication, they receive 2 points. If the user presses the final button again, then the third prompt appears (i.e., "You Have to Take Your Meds"). If the user follows this prompt and takes their medication, they then receive 1 point. If the user represses the final button without ever taking their medication, they get zero points, and the scene ends. Scores range from 0 to 6.
Visuospatial Attention	The user receives 2 point for each target item collected (6 target items). Also, up to 4 points are awarded for the speed of detecting the items. If the user attempts to collect one of the distractors, it counts as an error. Up to 4 points are awarded for the accuracy of detecting items. Scores range from 0 to 20.
Visual Attention	The user is awarded 1 point when a target poster is "spotted" and subtracted 1 point when a distractor poster is "spotted". Scores range from 0 to 16.
Auditory Attention	**Example:** if the user presses the trigger on the right controller to detect a target sound originating on the right side (i.e., controller and sound on the same side), then s/he gets 2 points. If the user presses the trigger on the right controller to detect a target sound originating on the left side (i.e., controller on the opposite side), s/he gains only 1 point. If the user responds to a distractor sound, irrelevant of its origin or the controller used to respond, 1 point is deducted. Scores range from 0 to 32.

Note: For all measures, higher scores indicate better performance.

Figure 1. VR-EAL Storyline: Scenes 3 - 12.

Scene 3 **Scene 3**



You need to draw your itinerary to the destinations

You are expected to respect the single-direction roads!

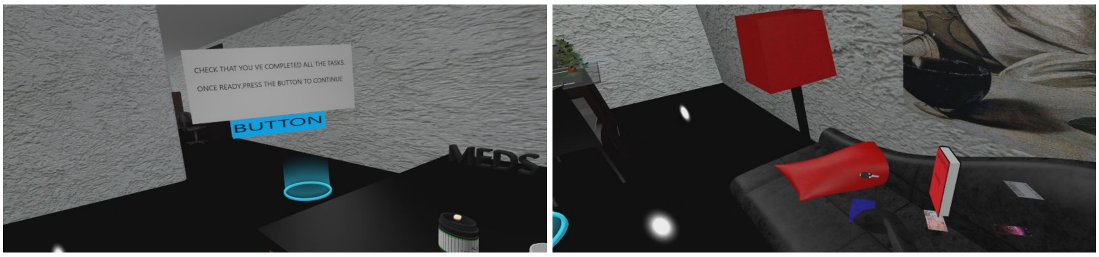
1st you need to go to the supermarket (as fast as possible)

2nd you need to go either to the bakery or to the library.
Though you need to visit both destinations!

Finally, after the bakery and the library you need to go back home!!
Once ready, press the button to start.
The destinations and the starting point (red road) will appear on the

Scene 6 **Scene 6**

Scene 6 **Scene 8**



CHECK THAT YOU'VE COMPLETED ALL THE TASKS.
ONCE READY, PRESS THE BUTTON TO CONTINUE

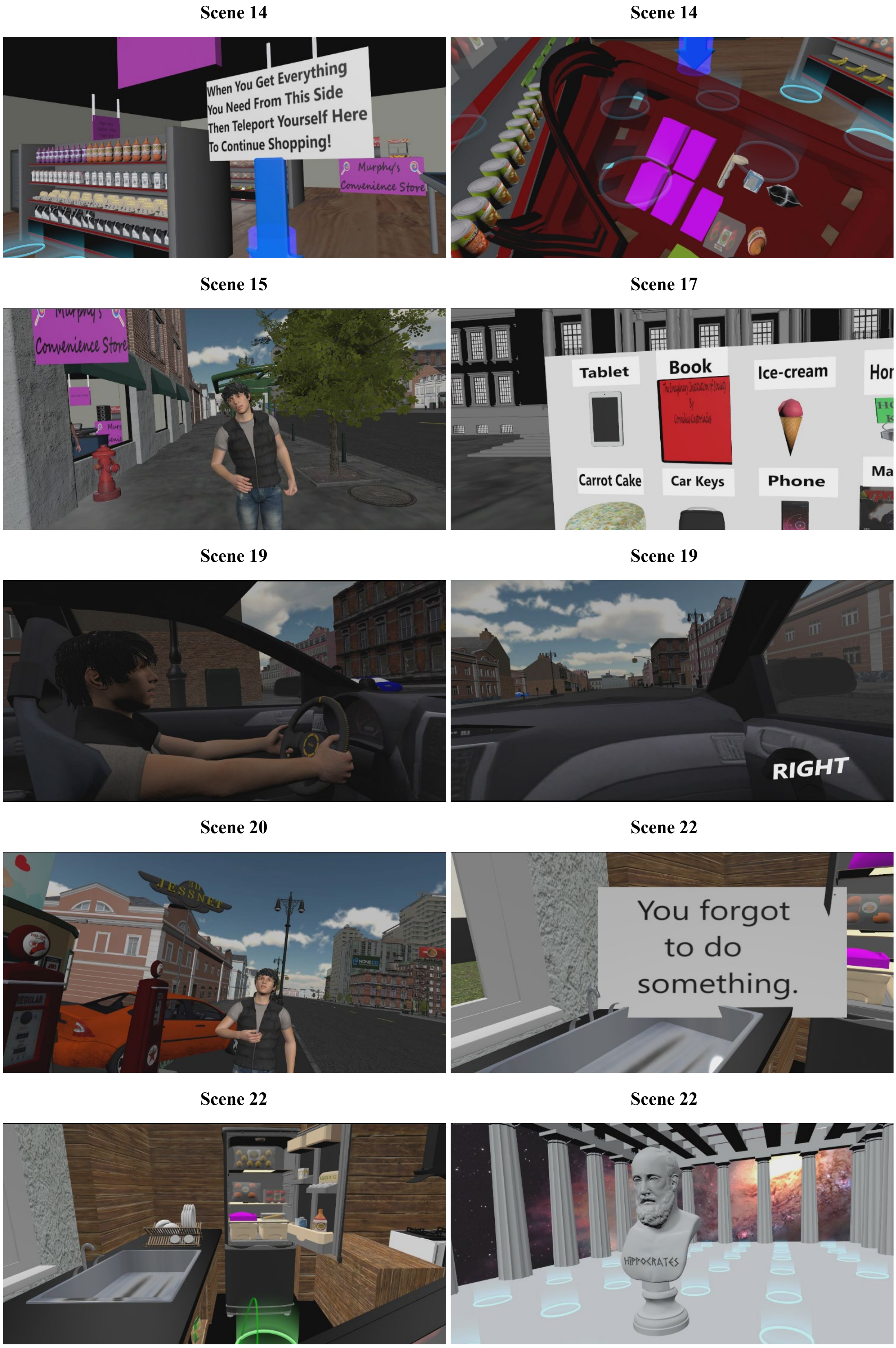
Scene 8 **Scene 10**

TAKE YOUR MEDS AFTER HAVING YOUR BREAKFAST.
TAKE THE CHOCOLATE PIE OUT OF THE OVEN & PLACE IT ON THE COUNTER BEFORE GOING OUT.
AT 10:00 AM - REMIND ALEX TO CALL ROSE AND INVITE HER TO THE DINNER.
AT 12:00 PM (NOON) GO TO THE BAKERY TO COLLECT THE CARROT CAKE.
RETURN THE BLACK BOOK WITH THE RED FRONT COVER TO THE LIBRARY
AT 1 PM REMEMBER TO TAKE YOUR LUNCH
WHEN YOU ARE BACK HOME GIVE THE THE FLAT TO ALEX.

Scene 12 **Scene 12**

Derived from Kourtesis et al., (2020b).

Figure 2. VR-EAL Storyline: Scenes 14 – 22



Scene 14 **Scene 14**

Scene 15 **Scene 17**

Scene 19 **Scene 19**

Scene 20 **Scene 22**

Scene 22 **Scene 22**

Derived from Kourtesis et al., (2020b).

Paper-and-Pencil Tests. Established ecologically valid paper-and-pencil test batteries in terms of both verisimilitude and veridicality were selected to match the equivalent VR-EAL tasks and examine their ecological and construct validity (i.e., CAMPROMPT, RBMT-III, BADS and TEA). Two additional neuropsychological tests that are ecologically valid in terms of only veridicality were also included to assess the validity of the VR-EAL's visuospatial attention and multitasking tasks. Tests of visuospatial attention and multitasking that are ecologically valid both in terms of verisimilitude and veridicality are not available in the literature.

Prospective Memory. The CAMPROMPT was administered to evaluate prospective memory using six prospective memory tasks (Wilson, 2005). Three tasks are event-based, and three are time-based. The participant is required to perform several distractor tasks (e.g., word-finder puzzles and general knowledge quizzes and questions) for 20 minutes, as well as remember to perform the prospective memory tasks (e.g., when the participant faces a question which includes the word "EastEnders", s/he needs to give a book to the examiner). The utilization of reminding strategies (e.g., taking notes) is permitted to aid the participant to remember when and how to perform the prospective memory tasks. The CAMPROMPT provides three scores: a total score (out of 36), an event-based score (out of 18), and a time-based score (out of 18).

Episodic Memory. Two subtests from the RBMT-III (Wilson *et al.*, 2008) were administered to assess episodic memory. The recall tasks were opted since they offer two scores (immediate recall, delayed recall), while the recognition tasks provide a score only for delayed recognition. The immediate and delayed story recall tasks were used to match the VR-EAL's immediate and delayed recognition tasks. The participant listens to a story from a newspaper read aloud by the examiner. The participant should recall the story immediately (immediate recall out of 21) and after approximately 20 minutes (delayed recall out of 21).

Executive Function: Planning. The Key Search task from the BADS (Wilson *et al.*, 1997) was utilized as a test of planning (Wilson, Evans, Emslie, Alderman, & Burgess, 1998). While the Key Search task assesses planning ability, it also involves other aspects of executive function (e.g., problem-solving, and monitoring of behavior; Wilson *et al.*, 1998). The participant should draw his or her route to find lost keys in a field. The quality of the route (e.g., whether it covers the whole field) and the time taken to draw it are considered in the scoring (max score = 16).

Executive functioning. The Color Trails Test (CTT; D'Elia, Satz, Uchiyama, & White, 1996) was administered to assess processing speed and executive functioning. CTT is a non-alphabetical adaptation (i.e., colors and numbers) of the Trail Making Test (Reitan, & Wolfson, 1993). CTT has two tasks (i.e., CTT-1 and CTT-2), where the participant must draw a line to connect consecutive numbers. In CTT-1, the numbers in the sequence are in a single color. Comparable to the TMT-A, CTT-1 assesses processing speed. In CCT-2, the numbers are displayed in two colors and the examinee alternates between the two colors for each number in the sequence. Comparable to the TMT-B, CTT-2 assesses task-switching, as well as inhibition and visual attention (D'Elia *et al.*, 1996). The CTT was chosen to assess the validity of the VR-EAL's multitasking task, and these aspects of executive functioning have been found central in everyday multitasking (Logie *et al.*, 2011). Furthermore, the time to complete in seconds is taken as the score for CTT-1 and CTT-2, and the difference between the two scores (i.e., CTT-2 minus CTT-1) is considered an index of executive function.

Selective visual attention. The Ruff 2 and 7 Selective Attention Test (RSAT; Ruff, Niemann, Allen, Farrow, & Wylie, 1992) was used to assess selective visual attention. The participant is asked to identify target numbers (i.e., 2s and 7s) and ignore the distractors (either numbers or letters) in the block. The examinee is required to implement two different strategies for each type of block; an automatic selection of 2s and 7s for the blocks with letter-distractors, and a controlled detection of 2s and 7s for the blocks with number-distractors. The RSAT produces two scores: a detection speed score (out of 80) and a detection accuracy score (out of 59). The scores consider the number of detected 2s and 7s, as well as, the number of misses and errors. The RSAT was opted to match the VR-EAL selective visuospatial attention task because it requires different scanning strategies, shifting of attention to another block, and considers the number of misses and mistakes.

Selective visual attention. The Map task from the TEA (Robertson *et al.*, 1994) was administered to assess selective visual attention (i.e., the ability to detect visual targets, while disregarding similar visual distractors). The participant should find as many as possible restaurant symbols (version A) or gas station symbols (version B) on a map of Philadelphia (USA) within two minutes. The total score out of 80 corresponds to the number of symbols detected overall, while one subscore corresponds to the number of symbols found in the first minute, and the other subscore refers to the number of symbols detected in the second minute.

Selective auditory attention. The Elevator Counting with Distraction task of the TEA (Robertson *et al.*, 1994) was administered, which measures auditory selective attention (i.e., the ability to select target sounds, while ignoring competitive auditory distractors). In each trial, the participant listens to different sounds (beeps), where s/he needs to count the number of normal pitched beeps (i.e., targets) and disregard the high pitched and low-pitched beeps (i.e., distractors). The total score is the number of correct responses across the 10 trials (max score = 10).

Questionnaires. Questionnaires were administered to examine the VR software quality and VRISE, gaming experience of the participants, as well as the verisimilitude and pleasantness of the tests. See Table 4 for a description of the questionnaires.

Table 4. Questionnaires' administration and scoring

Evaluation's Target	Administration and Scoring
VR software quality and VRISE.	The VRNQ was administered to assess the quality of the VR-EAL and the intensity of VRISE. The VRNQ is a 1–7 Likert scale questionnaire comprising 20 questions in total; 5 questions are pertinent to each of the 4 domains (i.e., user experience, game mechanics, in-game assistance and VRISE) (Kourtesis et al., 2019b). The assessed VRISE are nausea, dizziness, disorientation, fatigue, and instability. VRNQ produces a total score out of 140 and a subscore out of 35 for each domain. The parsimonious cut-offs of VRNQ were used to assess the suitability of VR-EAL (Kourtesis et al., 2020).
Gaming and VR experience	A survey questionnaire was administered to evaluate the gaming and VR experience of the participants (see Supplementary Material – Figure 1). The questionnaire (Likert scale 1-7) contains two questions regarding the weekly frequency of game playing and VR technology use, and two questions pertinent to the ability to play games and VR technologies use.
Verisimilitude and pleasantness	A comparison questionnaire (two versions, i.e., VR and paper-and-pencil) was administered to examine the participants' views on the pleasantness and ecological validity of the tests performed (see Supplementary Material – Figures 2 and 3). There were two separate versions of the comparison questionnaire with a Likert scale ranging from 1 to 7. There was one version for the VR-EAL tasks (see Supplementary Material – Figure 2), and another for the paper-and-pencil tests (see Supplementary Material – Figure 3). Both versions had the same two questions referring to the level of enjoyment (e.g., 1-highly unpleasant, 7-highly pleasant) and verisimilitude (e.g., 1-totally different from the tasks in daily life, 7-nearly identical to the tasks in daily life) of the tasks. For each version of the questionnaire, the maximum score was 14.

Procedure

Participants individually attended both the VR session and the paper-and-pencil session; the order was pseudorandomized across participants. In the VR session, participants participated in an induction session to introduce them to the HMD and controllers (i.e., HTC Vive and 6DoF wands-controllers) prior to immersion. After completion of VR-EAL, participants completed the VRNQ and the VR versions of the comparison questionnaire (i.e., to assess pleasantness and verisimilitude). During the paper-and-pencil session, participants completed the paper-and-pencil comparison questionnaires (i.e., pleasantness and verisimilitude) after each test. The duration of each session was timed using a stopwatch.

Statistical Analyses

A reliability analysis for the VR-EAL was conducted calculating Cronbach's alpha to inspect the internal consistency and reliability of the VR-EAL. A threshold of 0.70–1.00 for Cronbach's alpha was used, which indicates good (i.e., 0.70) to excellent (i.e., 1.00) internal consistency and reliability (Nunally & Bernstein, 1994).

The Bayesian factor (BF_{10}) was used for assessing statistical inference. The BF_{10} threshold ≥ 10 was set for statistical inference in all analyses, which indicates strong evidence in favor of the H1 (Marsman & Wagenmakers, 2017; Rouder & Morey, 2012; Wetzels & Wagenmakers, 2012) and corresponds to a p-value < 0.01 (e.g., $BF_{10} = 10$) (Bland, 2015; Cox & Donnelly, 2011; Held & Ott, 2018). BF_{10} is considered substantially more parsimonious than the p-value in evaluating the evidence against the H0 (Bland, 2015; Cox & Donnelly, 2011; Held & Ott, 2018), especially when evaluating the evidence of H1 against H0 in small sample sizes (Held

& Ott, 2018), as in the present study. Notably, BF_{10} allows evidence in either direction (i.e., towards H1 and H0), and its measurement of evidence is insensitive to the stopping rule, which substantially mitigates the issue of multiple comparisons and generates reliable and more generalizable results (Dienes, 2016; Marsman & Wagenmakers, 2017; Wagenmakers *et al.*, 2018).

Bayesian Pearson correlational analyses were conducted to examine associations between age, years of education, VR experience, gaming experience, and performance on the VR-EAL and paper-and-pencil tasks. Similarly, Bayesian Pearson correlational analyses were performed to assess construct validity for the entire VR-EAL and convergent validity between the VR-EAL tasks and the paper-and-pencil tasks. Furthermore, Bayesian paired samples t-tests were performed to investigate the differences between VR-EAL and paper-and-pencil tests in terms of verisimilitude, pleasantness, and administration time. Finally, a post hoc analyses for the achieved statistical power of the Bayesian Pearson's correlations and Bayesian paired samples t-tests were performed using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007; Faul, Erdfelder, Buchner, & Lang, 2009). All Bayesian analyses were performed using JASP (Version 0.8.1.2) (JASP Team, 2018).

Results

The descriptive statistics of the sample performing the VR-EAL, the paper-and-pencil tests and questionnaires are displayed in Table 5.

Table 5. Descriptive statistics for the VR-EAL, paper-and-pencil tests and questionnaires

	N	Mean (SD)	Range
Gaming Experience	41	6.12 (3.95)	2-13
VR Experience	41	3.29 (1.29)	2-6
Total Time VR-EAL (in minutes)	41	63.95 (7.88)	50-81
Total Time VR Session (in minutes)	41	73.95 (7.88)	60-91
Total Time Paper-Pencil Assessment (in minutes)	41	85.41 (3.97)	76-92
CAMPROMPT – Total Score (max = 36)	41	30.83 (3.49)	24-36
VR-EAL – PM Total Score (max = 48)	41	35.78 (4.73)	24-46
CAMPROMPT - Event Based (max = 18)	41	16.39 (1.63)	12-18
VR-EAL - Total Event Based (max = 24)	41	18.15 (3.26)	8-24
CAMPROMPT - Time Based (max = 18)	41	14.44 (2.66)	10-18
VR-EAL - Time Based (max = 18)	41	11.63 (3.10)	6-18
RBMT - Immediate Recall (max = 21)	41	14.93 (2.24)	10-18
VR-EAL - Immediate Recognition (max = 20)	41	15.51 (1.98)	10-18
RBMT - Delayed Recall (max = 21)	41	15.98 (2.61)	11-21
VR-EAL - Delayed Recognition (max = 20)	41	17.17 (2.42)	12-20
TEA – Map Total Score (max = 80)	41	70.32 (6.87)	52-82
VR-EAL - Selective Visual Attention Accuracy (max = 32)	41	22.98 (3.84)	17-30
RSAT – Accuracy (max = 59)	41	47.51 (7.14)	27-58
VR-EAL - Selective Visual Attention Speed (max = 32)	41	23.61 (3.69)	18-30
RSAT – Speed (max = 80)	41	57.78 (9.39)	33-74
VR-EAL - Selective Visuospatial Attention Total (max = 20)	41	12.00 (2.42)	4-15
VR-EAL - Selective Visuospatial Attention Speed (max = 16)	41	11.90 (1.50)	8-14
VR-EAL - Selective Visuospatial Attention Accuracy (max = 16)	41	12.10 (1.18)	8-13
TEA - Elevator Counting with Distraction (max = 10)	41	9.05 (1.05)	7-10
VR-EAL - Selective Auditory Attention (max = 32)	41	29.56 (3.66)	20-32
BADS – Key Search (max = 16)	41	14.20 (1.47)	10-16
VR-EAL – Planning (max = 19)	41	14.90 (1.51)	11-17
CTT – 1 (max = 80)	41	49.37 (8.65)	32-68
VR-EAL - Cooking Task (max = 16)	41	9.68 (2.57)	2-13
CTT – 2 (max = 80)	41	55.20 (9.94)	27-70

VR-EAL = Virtual Reality Everyday Assessment Lab; CAMPROMPT = Cambridge Prospective Memory Test; RBMT = Rivermead Behavioral Memory Test; TEA = Test of Everyday Attention; BADS = Behavioral Assessment of Dysexecutive Syndrome; CTT = Color Trails Test.

Correlations between demographics and performance

No significant correlations were found between age, education, VR experience, gaming experience, or performance on any of the paper-and-pencil tests or the VR-EAL tasks. The only significant correlations were observed between gaming experience and VR experience, VR experience and the VR session duration, gaming experience and the VR session duration, gaming experience and the duration of the paper-and-pencil testing session, and the duration of the VR session and the paper-and-pencil session (see Table 6).

Table 6. Bayesian correlations between users' experience and the sessions' durations.

Correlational Pairs	**r**	**BF_{10}**	**SP**
Gaming experience - VR experience	0.84^{***}	1.72e+10	~ 100%
VR experience - VR session duration	-0.60^{***}	690.55	99%
Gaming experience - VR session duration	-0.55^{***}	136.41	97%
Gaming experience - Paper-and-pencil session duration	-0.45^{***}	12.17	94%
VR session duration - Paper-and-pencil session duration	0.53^{***}	87.22	97%

The alternative hypothesis specifies that the correlation is positive. * $BF_{10} > 10$; ** $BF_{10} > 30$; *** $BF_{10} > 100$; r = Pearson's r; SP = Statistical Power at $\alpha < .05$;

Convergent and construct validity of the VR-EAL

The VR-EAL scores were significantly positively correlated with their equivalent scores on the paper-and-pencil tests (see Table 7). These results support the convergent validity of the VR-EAL tasks, as well as the construct validity of the VR-

EAL as an immersive VR neuropsychological battery. The reliability analysis demonstrated a Cronbach's $\alpha = 0.79$ for VR-EAL, which indicates good internal reliability (Nunally & Bernstein, 1994).

Table 7. Bayesian correlations between the VR-EAL and the paper-and-pencil tests

Paper-and-Pencil Scores	**VR-EAL Scores**	**r**	**BF_{10}**	**SP**
CAMPROMPT – Total	Total PM	0.82^{***}	3.20e+9	~ 100%
CAMPROMPT - Event Based	Event Based PM	0.73^{***}	3.97e+3	~ 100%
CAMPROMPT - Time Based	Time Based PM	0.67^{***}	2.61e+2	~ 100%

RBMT – Immediate Recall	Immediate Recognition	0.77^{***}	7.34e+7	~ 100%
RBMT – Delayed Recall	Delayed Recognition	0.82^{***}	3.90e+9	~ 100%
TEA – Map Total Score	Selective Visual Attention Accuracy	0.48^{**}	50.53	95%
TEA – Map Total Score	Selective Visual Attention Speed	0.46^{**}	34.99	93%
RSAT – Accuracy	Selective Visual Attention Accuracy	0.43^{*}	16.94	89%
RSAT – Accuracy	Selective Visuospatial Attention Total Score	0.61^{***}	2101	99%
RSAT – Speed	Selective Visuospatial Attention Speed	0.49^{**}	63.15	96%
RSAT – Accuracy	Selective Visuospatial Attention Accuracy	0.58^{***}	778.50	99%
TEA -Elevator Counting with Distraction	Selective Auditory Attention	0.70^{***}	8.91e+4	~ 100%
BADS – Key Search	Planning	0.80^{***}	4.65e+8	~ 100%
CTT – 1	Planning	0.47^{**}	41.74	94%
CTT – 2	Planning	0.51^{***}	109.73	97%
CTT – 1	Cooking Task	0.70^{***}	9.88e+4	~ 100%
CTT – 2	Cooking Task	0.80^{***}	8.75e+8	~ 100%
BADS – Key Search	Cooking Task	0.62^{***}	2.99e+3	99%

The alternative hypothesis specifies that the correlation is positive. * $BF_{10} > 10$; ** $BF_{10} > 30$; *** $BF_{10} > 100$; r = Pearson's r; SP = Statistical Power at $\alpha < .05$; VR-EAL = Virtual Reality Everyday Assessment Lab; CAMPROMPT = Cambridge Prospective Memory Test; RBMT = Rivermead Behavioral Memory Test; TEA = Test of Everyday Attention; RSAT = Ruff 2 and 7 Selective Attention Test; BADS = Behavioral Assessment of the Dysexecutive Syndrome; CTT = Color Trails Test

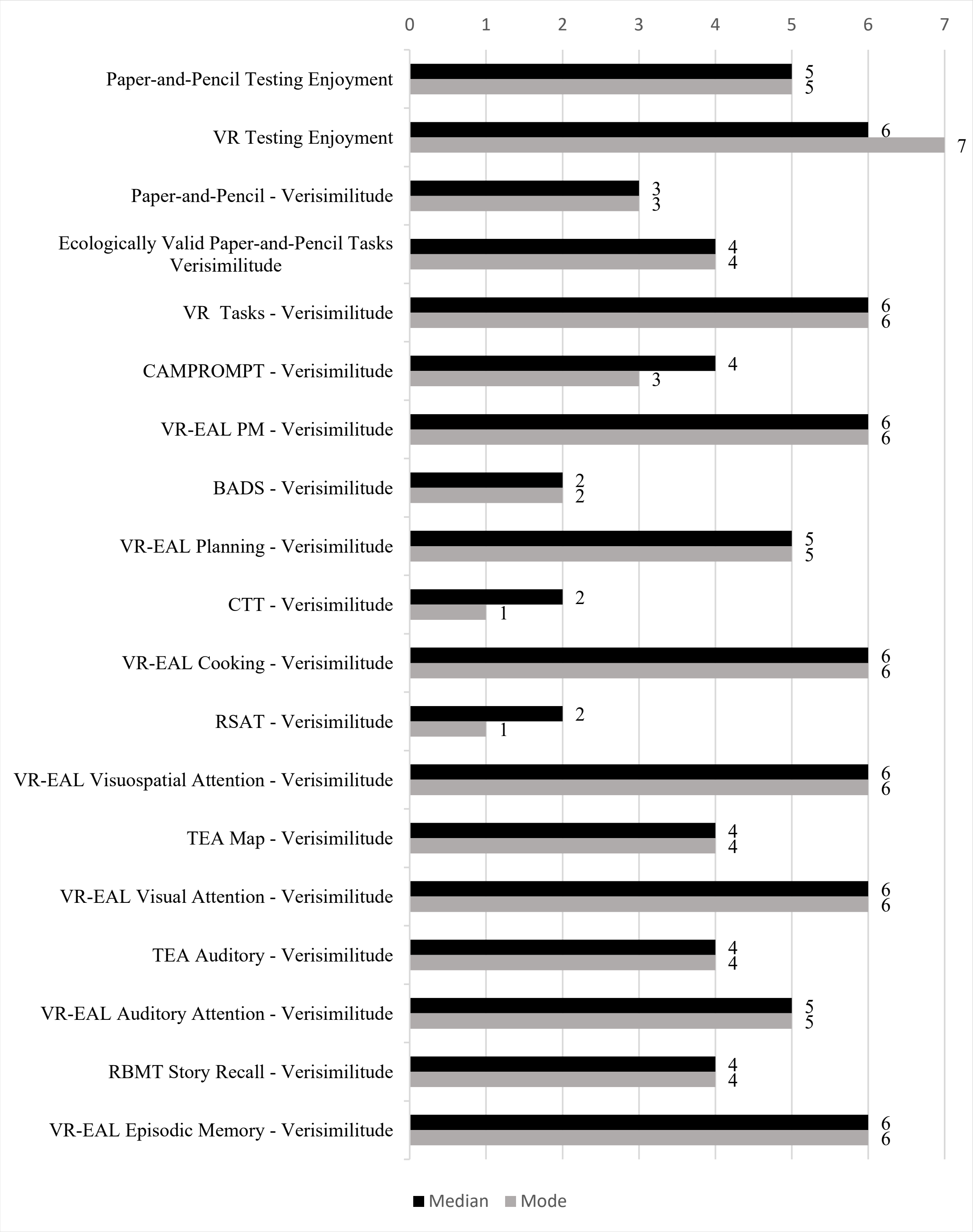
Quality of VR-EAL and VRISE: VRNQ

The median of the VRNQ total score for VR-EAL was 128, which is substantially above the parsimonious cut-off of 120 (maximum score = 140). The medians of the VRNQ domains (i.e., user experience, game mechanics, in-game assistance and VRISE) were between 31 and 33, again above their respective parsimonious cut-offs of 30 (maximum score = 35). Notably, the medians for all the individual VRISE items (i.e., nausea, dizziness, disorientation, fatigue, and instability) were 7 (i.e., absent feeling), except for fatigue, which was 6 (i.e., very mild feeling). No participant reported a VRISE subscore less than 5 (i.e., mild feeling).

Comparison of the testing experience between VR-EAL and paper-and-pencil tests

The median for enjoyment level was 6 (very pleasant) for the VR-EAL and 5 (pleasant) for the paper-and-pencil assessments (see Figure 3). The median for verisimilitude was 6 (i.e., very similar to everyday life) for VR-EAL, 4 (neither similar nor dissimilar to everyday life) for the ecologically validity tests, and 3 (dissimilar to everyday life) for the remaining paper-and-pencil tests (see Figure 3). The Bayesian t-tests demonstrated significant differences between the VR-EAL and paper-and-pencil tests, where the VR-EAL is rated significantly more pleasant and ecologically valid (i.e., verisimilitude) than the paper-and-pencil tests (see Table 8). In addition, the VR session was substantially shorter than the paper-and-pencil session (see Table 8).

Figure 3. Self-report verisimilitude and enjoyment of the VR-EAL and paper-and-pencil tests.



VR = Virtual Reality; CAMPROMPT = Cambridge Prospective Memory Test; PM = Prospective Memory; VR-EAL = Virtual Reality Everyday Assessment Lab; BADS = Behavioral Assessment of the Dysexecutive Syndrome; CTT = Color Trails Test; RSAT = Ruff 2 and 7 Selective Attention Test; TEA = Test of Everyday Attention; RBMT = Rivermead Behavioral Memory Test

Table 8. Comparison between administration time and participants' ratings of verisimilitude and enjoyment for the VR-EAL and paper-and-pencil tests

Paper-and-Pencil Test		**VR-EAL**	**BF_{10}**	**SP**
Total Administration Time	>	VR-Session Time	1.224e+11***	~ 100%
Testing Pleasantness	<	VR-Testing Pleasantness	188,842***	~ 100%
Total Verisimilitude	<	VR-EAL Verisimilitude	4.898e +15***	~ 100%
Ecologically Valid Tests/Tasks Verisimilitude	<	VR-EAL Verisimilitude	3.575e +13***	~ 100%
CAMPROMPT Verisimilitude	<	PM Verisimilitude	1.179e +9***	~ 100%
BADS Key Search Verisimilitude	<	Planning Verisimilitude	1.950e +13***	~ 100%
CTT Verisimilitude	<	Cooking Task Verisimilitude	6.849e +21***	~ 100%
RSAT Verisimilitude	<	Visuospatial Attention Verisimilitude	2.635e +13***	~ 100%
TEA Map Verisimilitude	<	Visual Attention Verisimilitude	3.774e +12***	~ 100%
TEA Elevator Counting with Distraction Verisimilitude	<	Auditory Attention Verisimilitude	4.36e +11***	~ 100%
RBMT Story Recall Verisimilitude	<	Episodic Memory Verisimilitude	1.244e +7***	~ 100%

* $BF_{10} > 10$; ** $BF_{10} > 30$; *** $BF_{10} > 100$; SP = Statistical Power at $\alpha < .05$; VR-EAL = Virtual Reality Everyday Assessment Lab; CAMPROMPT = Cambridge Prospective Memory Test; BADS = Behavioral Assessment of the Dysexecutive Syndrome; CTT = Color Trails Test; RSAT = Ruff 2 and 7 Selective Attention Test; TEA = Test of Everyday Attention; RBMT = Rivermead Behavioral Memory Test; PM = Prospective Memory; VR = Virtual Reality

Discussion

The VR-EAL was devised to assess cognitive functions (i.e., prospective memory, episodic memory, executive functions, and attentional processes) that are central to everyday functioning. Being an immersive VR research/clinical software, the VR-EAL aims to increase the likelihood that individuals' performance will replicate how they will act in real-life situations (Higginson *et al.*, 2000; Chaytor & Schmitter-Edgecombe, 2003; Phillips *et al.*, 2012; Rosenberg, 2015; Mlinac & Feng, 2016; Haines *et al.*, 2019). In the current study, we attempted to provide convergent, construct, and ecological validity for the VR-EAL tasks. Indeed, we demonstrated that all VR-EAL tasks significantly correlated with their corresponding ecologically valid paper-and-pencil tasks. The VR-EAL also showed good internal consistency, allowing implementation in clinical and research settings (Nunally & Bernstein, 1994). Therefore, the VR-EAL appears to be an effective, reliable, and ecologically valid tool for the assessment of everyday cognitive functioning, which can be used for clinical and research purposes. Importantly, the VR-EAL is a highly immersive and ergonomic VR neuropsychological battery; immersive VR provides a more ecological valid experience than non-immersive VR (Weech, Kenny, & Barnett-Cowan, 2019) and ergonomic interactions benefit non-gamers as their performance is comparable to gamers (Zaidi *et al.*, 2018).

Notably, the paper-and-pencil tests utilized in this study have been found to be ecologically valid in terms of both verisimilitude and veridicality (or veridicality only), evidencing their ability to predict everyday functioning. For example, the RBMT was highly accurate in predicting the everyday memory functionality of patients with traumatic brain injuries (TBI; Makatura, Lam, Leahy, Castillo, & Kalpakjian, 1999). Also, the RBMT has been strongly associated with occupational therapists' observations of general cognitive activities of daily living (ADL) in depressed and healthy older adults (Goldstein, McCue, Rogers, & Nussbaum, 1992). The RBMT, as well as the TEA, have been found to be the best predictors of general functional impairment in multiple sclerosis (MS) patients compared to traditional cognitive tests (Higginson *et al.*, 2000). The TEA was also successful in detecting cognitive aging effects in attentional processes in healthy old adults (Robertson *et al.*, 1994). The CAMPROMPT was a significant predictor of the occupational performance (e.g., returning to work, withdrawal, or compromised performance) in MS patients (Honan, Brown, & Batchelor, 2015) and the BADS was significantly associated with everyday executive skills (Evans, Chua, McKenna, & Wilson, 1997) and general cognitive performance (Norris & Tate, 2000) in neurological patients and healthy individuals. Equally, the Trail Making Test B (i.e., comparable to CTT-2) was a significant predictor of the everyday executive skills of neurological (Burgess *et al.*, 1998) and TBI patients (Chaytor, Schmitter-Edgecombe, & Burr, 2006). Lastly, the RSAT was found to be a key predictor of TBI patients' ability to return to professional or academic environments after rehabilitation (Ruff *et al.*, 1993).

Our findings regarding the convergent validity of the VR-EAL tasks with the corresponding paper-and-pencil tasks that have been established as predictors of real-world performance support the VR-EAL's ability to reflect performance outcomes in everyday life. However, the ecological validity of the VR-EAL would benefit from future work directly comparing the VR-EAL with true real-world functioning. For example, studies have shown that performance on real-world tasks (e.g., household chores) is significantly associated with self-ratings of instrumental activities of daily living (IADL) and independence questionnaires (Weakley, Weakley, & Schmitter-Edgecombe, 2019), which produce reliable and generalizable outcomes (Bottari, Dassa, Rainville, & Dutil, 2010; Bottari, Shun, Le Dorze, Gosselin, & Dawson, 2014). Thus, the predictive ability and/or veridicality of the VR-EAL could be further examined by investigating its relationship with real-world tasks and/or established IADL questionnaires in healthy older adults and/or clinical populations.

Nevertheless, considering that the verisimilitude approach may be more efficient than the veridicality approach in predicting everyday performance (Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006), our findings suggest that

VR-EAL's high verisimilitude is an advantage over other ecological valid tests. Previous studies examining the ecological validity of other VR neuropsychological tools have not considered users' perceptions of the task's verisimilitude (e.g., Canty *et al.*, 2014; Jansari *et al.*, 2014; Logie *et al.*, 2011; Rand *et al.*, 2009). Therefore, a further advantage of VR-EAL is that the participants rated it as more similar to the tasks that they perform in their daily life (i.e., more ecologically valid in terms of verisimilitude) than all tests in the paper-and-pencil neuropsychological battery and the group of well-established ecological valid tests with verisimilitude (i.e., CAMPROMPT test, RBMT-Story Recall, BADS-Key Search, TEA-Map, and TEA-Elevator Counting with Distraction). Furthermore, the VR-EAL tasks were individually compared to their corresponding paper-pencil test, where the results postulated that the VR-EAL tasks are significantly more ecologically valid in terms of verisimilitude than the equivalent paper-pencil tests. Also, as far as we are aware, our study is the first to compare the pleasantness of the testing experience between immersive VR and paper-and-pencil tests. Here, the full-version of the VR-EAL was also considered by the participants to be a more pleasant testing experience than the paper-and-pencil neuropsychological battery. Furthermore, the duration of the entire VR session (i.e., the induction and performance of VR-EAL) was considerably shorter than the administration time for the paper-and-pencil neuropsychological battery. Therefore, the VR-EAL emerges as substantially more enjoyable and ecologically valid testing experience with a significantly shorter administration time in comparison with the equivalent paper-and-pencil neuropsychological battery.

Age and education did not correlate with performance on the VR-EAL or the paper-and-pencil tests. While the paper-and-pencil scores were adjusted for age and education, the VR-EAL scores were not. Therefore, the VR-EAL may have the advantage that performance is not dependent on age or education. However, this needs to be further investigated in a larger and more diverse population, as the population of this study predominantly comprised younger adults aged 18 to 45 years with a relatively high level of education (i.e., 10-16 years).

Gaming experience strongly and positively associated with VR experience, indicating that gamers are also more experienced immersive VR users. Also, VR and gaming experience were both negatively correlated with the duration of the VR session, where more experienced gamers complete the assessment faster than non-gamers. Interestingly, however, gaming experience was also correlated with the duration of the paper-and-pencil session, indicating that gamers complete the paper-and-pencil assessment faster than non-gamers. Finally, the duration of the VR session was correlated significantly with that of the paper-and-pencil session, which also indicates that the speed of performing tasks affects the duration of both types of tasks (i.e., immersive VR and paper-and-pencil). Our findings are aligned with the relevant literature where gamers have been found to have enhanced perceptual processing speed (Anguera *et al.*, 2013; Dye, Green, & Bavelier, 2009; Kowal, Toth, Exton, & Campbell, 2018). However, in our sample, the gaming ability was not associated with the performance on the cognitive tests, indicating that gaming ability is not linked with an improved overall cognition, which is also in line with the relevant literature (Kowal *et al.*, 2018).

Another aim of the study was to provide immersive VR software for clinical and research use that has minimal VRISE, since adverse symptomology associated with VR can significantly decrease participants' reaction times and overall cognitive performance (Nalivaiko *et al.*, 2015; Nesbitt *et al.*, 2017; Mittelstaedt *et al.*, 2018). Albeit that the incidence of VRISE is more frequent in immersive VR, these symptoms are also highly frequent in non-immersive VR (Sharples, Cobb, Moody, & Wilson, 2008). However, the examination and report of VRISE has not been considered in non-immersive VR studies of neuropsychological tools for clinical and research purposes (e.g., Canty *et al.*, 2014; Jansari *et al.*, 2014; Logie *et al.*, 2011; Rand *et al.*, 2009). Similarly, the examination of VRISE is under-reported or not examined in immersive VR studies of neuropsychological tools (Kourtesis *et al.*, 2019a).

In contrast, the examination and report of VRISE was central in our endeavour to scrutinise the suitability of VR-EAL as a neuropsychological tool for research and clinical purposes. Our current findings replicate those of our previous work where VR-EAL did not induce VRISE in participants (Kourtesis *et al.*, 2020). In this study, VR-EAL exceeded the parsimonious cut-offs for the VRNQ scores (total score, user experience, game mechanics, in-game assistance, and VRISE). The outcomes of VRNQ hence postulate that VR-EAL is a suitable VR software for implementation in research and clinical settings, without inducing VRISE. On all VRISE items, except fatigue, there was an absence of adverse symptoms. Participants reported only very mild feelings of fatigue albeit that this was an expected outcome since the duration of VR-EAL was around 60 minutes. However, fatigue was equally present during the paper-and-pencil session (80 minutes).

This study also has some limitations. The sample was moderately small (N = 41), though, every statistical analysis displayed a substantially robust statistical power (>90%). Moreover, as the current study is the first to provide validity for the VR-EAL, it was only administered to younger but not older adults. Yet, the eventual aim is to use the VR-EAL to assess cognitive impairments in healthy aging and dementias (Anderson & Craik, 2017) or attention-deficit/hyperactivity disorder and autism (Karalunas *et al.*, 2018). Future work should examine the performance and experiences of different clinical populations performing the VR-EAL to provide further evidence for the clinical utility of VR-EAL for assessing everyday cognitive functioning.

In summary, this study provides evidence supporting the validation of VR-EAL as an effective neuropsychological tool with enhanced ecological validity for the assessment of everyday cognitive functioning. In addition, the VR-EAL does not seem to induce VRISE (i.e., cybersickness). Therefore, our preliminary findings support the VR-EAL as an immersive VR assessment tool that has the potential to be implemented in both research and clinical settings in the future.

Chapter 5 provided evidence for the psychometric properties of the VR-EAL such as its content and construct validity. Also, this chapter demonstrated that the VR-EAL offers a more pleasant and ecologically valid neuropsychological assessment of everyday cognitive functions with a substantially shorter administration time against the equivalent paper-and-pencil neuropsychological battery. Also, Chapter 5 replicated the findings of Chapter 4 regarding the mitigation of VRISE by the VR-EAL. The findings of Chapter 5 hence demonstrated the utility of the VR-EAL as a neuropsychological assessment. However, as discussed in Chapter 4 the VR-EAL was also developed in line with the principal theoretical frameworks of prospective memory. Therefore, the capability of the VR-EAL to contribute to the understanding of the everyday prospective memory should be examined.

Chapter 6: An Ecologically Valid Examination of the Theoretical Framework of Prospective Memory: The Facilitation of Real-World Prospective Memory by Attention, Memory, and Executive Function Processes.

Chapter 5 demonstrated the psychometric properties and utility of the VR-EAL. Nevertheless, as discussed in Chapter 4, the VR-EAL was designed in line with the preparatory attentional and memory processes (PAM) and mutiprocess theories, which are the predominant theoretical frameworks of prospective memory. The VR-EAL assesses the whole spectrum of prospective memory (i.e., focal event-based, non-focal event-based, and time-based task), as well as other cognitive processes pertaining to attention, episodic memory, and executive functions, which are associated with prospective memory. In Chapter 6, using the VR-EAL, an ecological valid examination of the principal theoretical frameworks of prospective memory will be conducted, as well as an exploration of the cognitive processes which facilitate everyday prospective memory functioning.

This chapter has been submitted for publication in the Memory & Cognition journal: Kourtesis, P., Collina, S., Doumas, L. A., & MacPherson, S. E. (submitted). An ecologically valid examination of the theoretical framework of prospective memory: the facilitation of real-world prospective memory by attention, memory, and executive function processes. *Memory & Cognition*.

An ecologically valid examination of the theoretical framework of prospective memory: the facilitation of real-world prospective memory by attention, memory, and executive function processes.

Panagiotis Kourtesis[a,b,c,d]*, Simona Collina[c,d], Leonidas A.A. Doumas[b], and Sarah E. MacPherson[a,b]

[a] *Human Cognitive Neuroscience, Department of Psychology, University of Edinburgh, Edinburgh, UK;*

[b] *Department of Psychology, University of Edinburgh, Edinburgh, UK;*

[c] *Lab of Experimental Psychology, Suor Orsola Benincasa University of Naples, Naples, Italy;*

[d] *Interdepartmental Centre for Planning and Research "Scienza Nuova", Suor Orsola Benincasa University of Naples, Naples, Italy*

* Panagiotis Kourtesis, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, Scotland, United Kingdom Email: pkourtes@exseed.ed.ac.uk

Abstract

The main theories explaining prospective memory (PM) functioning are the preparatory attentional and memory processes (PAM) and the multiprocess theories. These theories mainly explain focal and non-focal event-based PM but not time-based PM or the role of executive functions in PM. Notably, the evidence supporting the PAM and multiprocess theories mainly derive from non-ecological laboratory tasks. Ecologically valid neuropsychological tasks resemble the complexity and cognitive demands of everyday tasks and allow a generalization of the findings to real-life cognitive functioning. The Virtual Reality Everyday Assessment Lab (VR-EAL), an immersive virtual reality neuropsychological battery with enhanced ecological validity, was implemented to assess everyday PM (i.e., focal and non-focal event-based, and time-based), episodic memory, attentional processes, and executive functions. Bayesian statistics examined the differences between PM tasks and the predictors of PM scores. The results revealed that the type of PM task does not play a significant role in everyday PM functioning, but instead the length of delay between encoding and retrieving the PM intention plays a central role. Support for the PAM and multiprocess frameworks was found in non-focal and focal event-based tasks respectively. However, the findings, inferring a dynamic interplay between automatic and intentional monitoring and retrieval processes, agree with the inclusive approach of the multiprocess framework. Also, the role of executive functions appears crucial in everyday PM. Finally, everyday PM is predominantly facilitated by episodic memory, visuospatial attention, auditory attention, and executive functions.

Keywords: *event-based prospective memory; time-based prospective memory; multiprocess framework; preparatory attentional and memory processes; executive functions*

Background & Rationale

Prospective memory (PM) involves the ability to remember to initiate an action in the future (Brandimonte, Einstein, & McDaniel, 2014). The PM action may be associated with either a certain event (event-based e.g., when you see Alex, give him this card) or a specific time (time-based e.g., at 3 pm call John; Einstein & McDaniel, 1996). The two predominant theoretical models of PM are the preparatory attentional and memory (PAM) theory and the multiprocess theory (Anderson, McDaniel, & Einstein, 2017). The PAM theory suggests that an efficient PM ability stems from a constant top-down monitoring of environmental and internal cues which allow the individual to recall the intended action and initiate it (Smith, 2003; Smith, Hunt, McVay, & McConnell, 2007). For example, if an individual wants to buy a loaf on their way home from work, they engage attentional processes to monitor their environment for cues (e.g., a bakery sign) that will remind them of their intention to buy a loaf. The multiprocess theory is complementary to the PAM theory (McDaniel & Einstein, 2000; McDaniel & Einstein, 2007). In addition to PAM's top-down monitoring, the multiprocess theory postulates that effective performance on PM tasks is also facilitated by bottom-up spontaneous retrieval (McDaniel & Einstein, 2000; McDaniel & Einstein, 2007). In the previous example, when the individual is not engaging their attentional processes, they may see a billboard displaying pastry products, which then triggers the retrieval of the intention to buy a loaf. Hence, when retrieving the intention to perform a PM task, the PAM theory suggests there is an intentional retrieval process through the constant monitoring for cues, while the multiprocess theory adds a reflexive associative retrieval process by passively detecting environmental and internal cues (McDaniel, Umanath, Einstein, & Waldum, 2015).

There are, however, further differences between the PAM and multiprocess theories. One difference is the salience of the internal or environmental cue that triggers the retrieval of the PM intention. For example, while both the PAM and multiprocess theories state that preparatory attentional processes are required when the cue is non-focal (e.g., a sign amongst several signs for the nearby shops and buildings), when the cue is focal (e.g., a large and easily detectable billboard on the side of the road), only the PAM theory predicts that preparatory attentional processes are necessary for high levels of PM performance (Einstein, Smith, McDaniel, & Shaw, 1997; Einstein, McDaniel, Manzi, Cochran, & Baker, 2000; McDaniel *et al.*, 2015; Smith, 2003). Another difference between the theories is that individuals perform substantially better on focal PM tasks than non-focal PM tasks, which indicates the difference in difficulty between the intentional and reflexive associative retrieval of the intention for each type of PM task (Anderson *et al.*, 2017; McDaniel *et al.*, 2015; Mullet, Scullin, Hess, Scullin, Arnold, & Einstein, 2013; Scullin, McDaniel, Shelton, & Lee, 2010). Lastly, the PAM theory postulates that preparatory processes pertaining to memory (e.g., encoding and maintaining the PM intention) are equally important as attentional processes for the successful retrieval of intentions (Smith, 2003; Smith *et al.*, 2007). The multiprocess theory emphasizes that greater overlap between the processing required by the ongoing task (i.e., non-PM tasks) and the processing required by the PM cue assists in better encoding and retrieval of the PM intention (Anderson *et al.*, 2017; McDaniel & Einstein, 2000; McDaniel & Einstein, 2007; McDaniel *et al.*, 2015).

The PAM theory argues that the retrieval of the intention is always non-automatic to some degree (Smith, 2003; Smith *et al.*, 2007). The multiprocess theory agrees that there are non-automatic retrieval processes but also postulates an interplay between automatic and non-automatic retrieval processes (McDaniel & Einstein, 2000; McDaniel & Einstein, 2007). McDaniel and Einstein (2000, 2007) argued that it is unlikely that individuals in the real-world are constantly monitoring for cues; that would be cognitively unsustainable. PM functioning in everyday life is likely facilitated by both reflexive associative retrieval and intentional monitoring and retrieval (McDaniel & Einstein, 2000; McDaniel & Einstein, 2007).

More recently, McDaniel and collaborators (2015) provided methodological considerations for studies examining the multiprocess theory. These suggestions include the use of event-based PM tasks and a suspended intention paradigm (e.g., to instruct participants that there is a PM task but they need to perform another task now and the PM task later) (McDaniel *et al.*, 2015). Also, the event-based tasks need to have as many characteristics as possible from the following list: a) a single focal cue; b) minimal cues prompting individuals to focus on the PM task(s) rather than the ongoing tasks; c) emphasize the importance of the ongoing task(s) and minimize the importance of the PM task(s); d) use long-lasting ongoing task(s), while the occurrence of PM cue(s) should be delayed and minimized; e) instruct the participants that they may perform the PM task at any point after seeing the PM cue (McDaniel *et al.*, 2015).

Since PM functioning in everyday life consists of both event- and time-based tasks (Einstein & McDaniel, 1996), including only event-based tasks in the assessment of PM does not allow for generalization of the results to PM in the real world (e.g., Einstein *et al.*, 1997, 2000; Mullet *et al.*, 2013; Scullin *et al.*, 2010; Smith, 2003; Smith *et al.*, 2007). The literature suggests that time-based tasks are more cognitively demanding than event-based tasks as individuals frequently perform more poorly on time-based tasks compared to event-based tasks (Einstein & McDaniel, 1996). Also, PM is supported by attention demanding processes that are strongly associated with executive functioning, which monitor the environment for prospective cues (e.g., Smith & Bayen, 2004), inhibit performance of the ongoing task, and switch to the PM intention at the correct time (Marsh, Hicks, & Watson, 2002; West, 2011). While there is ample evidence for a robust relationship between PM and executive functions (e.g., Azzopardi, Auffray, & Kermarrec, 2017; Gonneaud *et al.*, 2011; Schnitzspahn, Stahl, Zeintl, Kaller, & Kliegel, 2013; Zuber, Kliegel, & Ihle, 2016; Zuber *et al.*, 2019), the executive functions that underlie and predict performance on these different PM tasks in everyday life seems somewhat unclear. The most adopted approach to study the relationship between PM and executive functions is the three-function model (Miyake *et al.*, 2000). This model comprehends three distinct executive functioning processes: 1) updating (i.e., controlling the stored information in working memory); 2) inhibition (i.e., refraining from responding prematurely and resisting distractions); and 3) shifting (i.e., shifting attention between different tasks). These executive functioning processes have been found to facilitate event-based and time-based PM functioning (Azzopardi *et al.*, 2017; Gonneaud *et al.*, 2011; Schnitzspahn *et al.*, 2013; Zuber *et al.*, 2016, 2019). Therefore, the diverse types of PM tasks (i.e., focal event-based, non-focal event-based, and time-based) and executive functioning processes should be included in paradigms that can generalize to everyday PM functioning.

Another criticism is the ecological validity of the task(s) that assess PM functioning. Ecologically valid neuropsychological tasks resemble the complexity and cognitive demands of corresponding everyday tasks (Franzen & Wilhelm, 1996; Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). Implementing ecologically valid tasks allows one to generalize the findings to everyday functioning (Chaytor & Schmitter-Edgecombe, 2003; Haines *et al.*, 2019; Higginson, Arnett, & Voss, 2000; Mlinac & Feng, 2016; Parsons, 2015; Phillips, Henry, & Martin, 2008; Rand *et al.*, 2009; Rosenberg, 2015). The existing findings pertaining to the PAM and multiprocess theories tend to derive from laboratory experiments (e.g., Einstein *et al.*, 1997, 2000; Mullet *et al.*, 2013; Scullin *et al.*, 2010; Smith, 2003; Smith *et al.*, 2007). These laboratory tasks incorporate simple, static stimuli within a highly controlled environment and do not resemble the complexity of real-life situations (Parsons, 2015; Rand, Rukan, Weiss, & Katz, 2009). PM studies adopting experimental tasks that diverge from being ecologically valid may provide discrepant results compared to ecologically valid research paradigms (Marsh, Hicks, & Landau, 1998). Lastly, ecologically valid PM research paradigms may elucidate processes that facilitate everyday PM functioning such as attentional switching, sustained attention, retrospective memory, and metamemory (e.g., knowledge of one's own strategies to encode, redefine, and consolidate PM intentions; Marsh *et al.*, 1998). The implementation of an ecologically valid research paradigm therefore is required to scrutinize the complex structure of everyday PM functioning.

Ecologically valid research paradigms have predominantly involved assessments in real-life settings such as performing errands in a shopping centre or a pedestrianized street (e.g., Garden, Phillips, & MacPherson, 2001; Marsh *et al.*, 1998; Shallice & Burgess, 1991). However, real-world tasks suffer from several limitations such as an inability to be standardized for use in other clinics or laboratories, reduced feasibility with certain neurological populations (e.g., schizophrenia patients), augmented time, fiscal, procedural and bureaucratic demands, as well as diminished experimental control over external factors (Elkind, Rubin, Rosenthal, Skoff, & Prather, 2001; Logie, Trawley, & Law, 2011; Parsons, 2015; Rand *et al.*, 2009). Another approach for achieving ecological validity pertains to the implementation of technological mediums such as recordings of real-world locations and non-immersive virtual environments (Farrimond, Knight, & Titov, 2006; Logie *et al.*, 2011; McGeorge *et al.*, 2001; Paraskevaides *et al.*, 2010; Rand *et al.*, 2009), which are able to simulate real-life tasks, are cost-effective, require less administration time, enable increased experimental control, and can be implemented in other clinical or research settings (Parsons, McMahan, & Kane, 2018; Werner & Korczyn, 2012; Zygouris & Tsolaki, 2015). However, non-immersive virtual reality (VR) tests can be challenging for individuals without gaming backgrounds (e.g., familiarization with using a keyboard and mouse to interact with the virtual environment; Zaidi, Duthie, Carr, & Maksoud, 2018) and they are substantially less ecologically valid than immersive VR tests (Parsons *et al.*, 2018). In contrast, immersive VR tests benefit from the same advantages as non-immersive ones, while they facilitate an enhanced ecologically valid testing environment (Bohil, Alicea, & Biocca, 2011; Parsons, 2015; Rizzo, Schultheis, Kerns, & Mateer, 2004; Teo *et al.*, 2016). Also, the first-person view and ergonomic interactions that are proximal to real-life actions in immersive VR environments significantly mitigate the differences in performance between gamers and non-gamers (Zaidi *et al.*, 2018). Consequently, an immersive VR research paradigm would be the most efficient approach to study everyday PM functioning.

We have recently developed the Virtual Reality Everyday Assessment Lab (VR-EAL), which is an immersive VR neuropsychological battery assessing everyday cognitive functions such as PM (i.e., event-based and time-based), episodic memory (i.e., immediate and delayed recognition), attentional processes (i.e., visual, visuospatial, and auditory), and executive functioning (i.e., planning and task-shifting; Kourtesis, Korre, Collina, Doumas, & MacPherson, 2020b). The convergent, construct, and ecological validity of the VR-EAL has been maintained against established ecologically valid paper-and-pencil tests such as the Cambridge Prospective Memory Test (Wilson *et al.*, 2005), the Rivermead Behavioral Memory Test–III (Wilson, Cockburn, & Baddeley, 2008), the Test of Everyday Attention (Robertson, Ward, Ridgeway, & Nimmo-Smith, 1996), and the Behavioral Assessment of Dysexecutive Syndrome (Wilson, Alderman, Burgess, Emslie, & Evans, 1997; Kourtesis, Collina, Doumas, & MacPherson, 2020a). Notably, the VR-EAL does not induce differences in performance between gamers and non-gamers, and the VR-EAL tasks were rated by participants as substantially more similar to everyday tasks than the equivalent ecologically valid paper-and-pencil tasks (Kourtesis *et al.*, 2020a). Hence, the VR-EAL appears to resemble the complexity and cognitive demands of real-life tasks.

The VR-EAL includes PM tasks which are consistent with both the PAM and multiprocess theories, while assessing both event-based and time-based PM (Kourtesis *et al.*, 2020b). The VR-EAL incorporates five event-based tasks and four time-based tasks. The event-based tasks consist of two non-focal event-based tasks, two focal event-based tasks, and one misleading task (i.e., prompting the examinee to perform a task when there is no task to perform). Also, the focal and non-focal event-based tasks are further segregated into tasks performed at an early or late stage of the scenario. Similarly, the time-based tasks are divided into early (two tasks) and late tasks (two tasks), where one of the two late tasks is a misleading task. The segregation of the PM tasks into early and late tasks allows us to investigate the effect of the length of the delay between encoding the PM intention and performing the PM action. Moreover, the assessment of PM in the VR-EAL is aligned with the methodological suggestions by McDaniel and collaborators (2015). There are only two focal event-based tasks within a scenario of approximately 70 minutes. The focal event-based tasks have a single focal cue each and there are no other cues driving the participant's attention to the PM task. Participants are instructed that the PM tasks need to be performed later in the scenario (i.e., anytime between 15 and 60 minutes after forming a PM intention). Finally, the importance of PM and non-PM tasks is equally highlighted, while non-PM tasks appear of high importance (e.g., planning the itinerary in the city, preparing breakfast, and buying groceries from the supermarket).

The non-PM VR-EAL tasks do not serve simply as distractors but are cognitively demanding tasks which assess cognitive functions associated with PM functioning. In line with the PAM theory, the VR-EAL assesses attentional (i.e., visual, visuospatial, and auditory attention) and memory (i.e., immediate and delayed recognition) processes, which are consistently found to contribute to PM functioning (Cona, Scarpazza, Sartori, Moscovitch, & Bisiacchi, 2015). Also, the VR-EAL assesses planning and task-shifting abilities. In line with the three-function model, the cooking task (i.e., multitasking/task-switching) pertains to updating the stored information in working memory, shifting attention between different tasks (i.e., switching), as well as inhibiting premature responses or actions prompted by distractions (i.e., inhibition; Craik & Bialystok, 2006; Doherty, Barker, Denniss, Jalil, & Beer, 2015; Zuber *et al.*, 2019). Similarly, the planning task involves the control of information stored in working memory (i.e., updating; Zuber *et al.*, 2019). Both planning and task-shifting have also been found to contribute to PM functioning (e.g., Azzopardi *et al.*, 2017; Gonneaud *et al.*, 2011; Schnitzspahn *et al.*, 2013; Zuber *et al.*, 2016, 2019).

The present study is the first to adopt an ecologically valid paradigm using immersive VR to examine real-life PM and cognition. Also, our study is the first to assess focal event-based, non-focal event-based, and time-based PM, as well as attentional, episodic memory, and executive functioning processes pertinent to PM functioning in the same sample to investigate whether the observed differences in laboratory paradigms will be replicated in our ecologically valid paradigm. Similarly, we compared performance on early and late stage PM tasks to examine the effect of the delay from forming the PM intention (i.e., encoding) until performing the PM action (i.e., retrieval) on PM functioning. Finally, we explored the cognitive functions facilitating everyday PM functioning.

Methods

Participants

Participants were recruited via social media and the internal mailing list of the University of Edinburgh and were the same cohort as in Kourtesis et al. (2020a).

Forty-one participants (21 females) with a mean age of 29.15 years (SD = 5.80, range = 18-45) and mean education of 13.80 years (SD = 2.36, range = 10-16) were recruited. The study was approved by the Philosophy, Psychology and Language Sciences Research Ethics Committee of the University of Edinburgh. Written informed consent was obtained from each participant. All participants received verbal and written instructions regarding the procedures, possible adverse effects of immersive VR (e.g., cybersickness), utilization of the data, and general aims of the study.

Materials

Hardware.

An HTC Vive HMD with two lighthouse stations for motion tracking and two HTC Vive wands with six degrees of freedom (6DoF) for navigation and interactions within the virtual environment were implemented in accordance with our previously published technological recommendations for immersive VR research (Kourtesis, Collina, Doumas, & MacPherson, 2019). The spatialized (bi-aural) audio was facilitated by a pair of Senhai Kotion Each G9000 headphones. The size of the VR area was $5m^2$, which provides an adequate space for immersion and naturalistic interaction within virtual environments (Borrego, Latorre, Alcañiz, & Llorens, 2018). The HMD was connected to a laptop with an Intel Core i7 7700HQ 2.80GHz processor, 16 GB RAM, a 4095MB NVIDIA GeForce GTX 1070 graphics card, a 931 GB TOSHIBA MQ01ABD100 (SATA) hard disk, and Realtek High Definition Audio.

VR-EAL.

VR-EAL assesses everyday cognitive functions such as PM, episodic memory (i.e., immediate and delayed recognition), executive functioning (i.e., planning, multitasking) and selective visual, visuospatial and auditory (bi-aural) attention within a realistic immersive VR scenario lasting around 70 minutes (Kourtesis *et al.*, 2020a). See Table 1 and Figures 1 and 2 for a summary of the VR-EAL tasks assessing each cognitive ability. The VR-EAL tasks' descriptions are presented in Table 2, and VR-EAL tasks' scoring is shown in Table 3. Lastly, see Table 4 for a description of each type of VR-EAL PM task and the comparisons between them. For a full description of the VR-EAL's scenarios, tasks, and scoring, see Kourtesis *et al.* (2020). Also, a brief video recording of the VR-EAL may be accessed at this hyperlink: https://www.youtube.com/watch?v=IHEIvS37Xy8&t= .

Table 1. VR-EAL tasks and score ranges

Scene	Cognitive Function	Task	Score Ranges
3	Prospective memory	Write down the notes for the errands.	0 – 6
3	Immediate recognition	Recognising items on the shopping list.	0 – 20
3	Planning	Drawing the route to be taken.	0 – 19
6	Multitasking	Cooking task (preparing breakfast).	0 – 16
6	Prospective memory – event based	Take medication after breakfast.	0 – 6
8	Selective visuospatial attention	Collect items from the living room.	0 – 20
8	Prospective memory – event based	Take the chocolate pie out of the oven.	0 – 6
10	Prospective memory – time based	Call Rose at 10 am.	0 – 6
12	Selective visual attention	Find posters on both sides of the road.	0 – 16
14	Delayed recognition	Recognising items from the shopping list.	0 – 20
15	Prospective memory – time based	Collect the carrot cake from the bakery at 12 pm.	0 – 6
16	Prospective memory – event based	False prompt before going to the library.	-6 – 0
17	Prospective memory – event based	Return the red book to the library.	0 – 6
19	Selective auditory attention	Detect sounds from both sides of the road.	0 – 32
20	Prospective memory – time based	False prompt before going back home.	-6 – 0
21	Prospective memory – event based	Back home, give the extra pair of keys to Alex.	0 – 6
22	Prospective memory – time based	Take the medication at 1pm.	0 – 6

**The tasks are presented in the same order as they are performed within the scenario. The table derives from Kourtesis et al., (2020a).*

Table 2. VR-EAL tasks' descriptions

Type	Description
Prospective Memory	VR-EAL considers both event-based and time-based PM tasks. The PM tasks are divided into five event-based tasks and four time-based tasks. In the event-based tasks, the participant should remember to perform a PM action when a specific event occurs (e.g., take medicines after breakfast). In the time-based tasks, the examinee should remember to perform a planned action at a specific time (e.g., call Rose at 12 pm).
Episodic Memory	Both immediate and delayed episodic memory are assessed. Firstly, the participant needs to memorize a shopping list which is presented audio-visually. Immediately after the presentation of the list, the participant is presented with 30 items and should visually recognize and select the 10 items from the shopping list (immediate recognition). Participants are then expected to choose the items from the list when they arrive at the supermarket approximately 20 minutes later (delayed recognition). The examinee should choose the ten target items (i.e., create the shopping list) from an extensive array of items, which also contains five qualitative distractors (e.g., semi-skimmed milk versus skimmed milk), five quantitative distractors (e.g., 1 kg potatoes versus 2 kg potatoes), and ten false items (e.g., bread, bananas etc.).
Multitasking/ Task-Shifting	Multitasking is examined using a cooking task, where the participant should prepare and serve his or her breakfast (e.g., sausages, omelet, and a cup of tea/coffee) and place a chocolate pie in the oven. Scoring relies on the animations from each game object (i.e., the omelet and the sausages). At the beginning of the animation, both items have a reddish (raw) color which gradually turns to either a yellowish (omelet) or brownish (sausages) color, and finally both turn to black (burnt).
Planning	Planning ability is assessed by asking participants to draw his or her route around the city (e.g., visiting the bakery, supermarket, library, and returning home) on a 3D interactive board. The road system comprises 23 street units.
Visuospatial attention	Visuospatial attention is assessed by asking the participant to find and collect 6 specific items (i.e., a mobile phone, a £50 note, a library card, the flat keys, a red book, and car keys) in the living-room. A reminder of these items remains on the wall (i.e., the items are displayed as 3D objects with labels). However, there are also distractors (i.e., magazines, books, a remote control, a notebook, a pencil, a chessboard, and a bottle of wine) in the room.
Visual attention	Visual attention is measured while the participant is seated as a passenger in a car next to a driver. The participant should identify all the targets (i.e., 16 posters of a radio station) on both sides of the road, while s/he needs to avoid any distractors (i.e., 8 posters that are a different shape and 8 posters with a different background color).
Auditory attention.	Auditory attention is also examined while the participant is seated as a passenger next to a driver. The participant should detect all the target sounds (i.e., 16 bell sounds) presented on both sides of the road, while avoiding the distractor sounds (i.e., 8 high pitch bells, and 8 dongs).

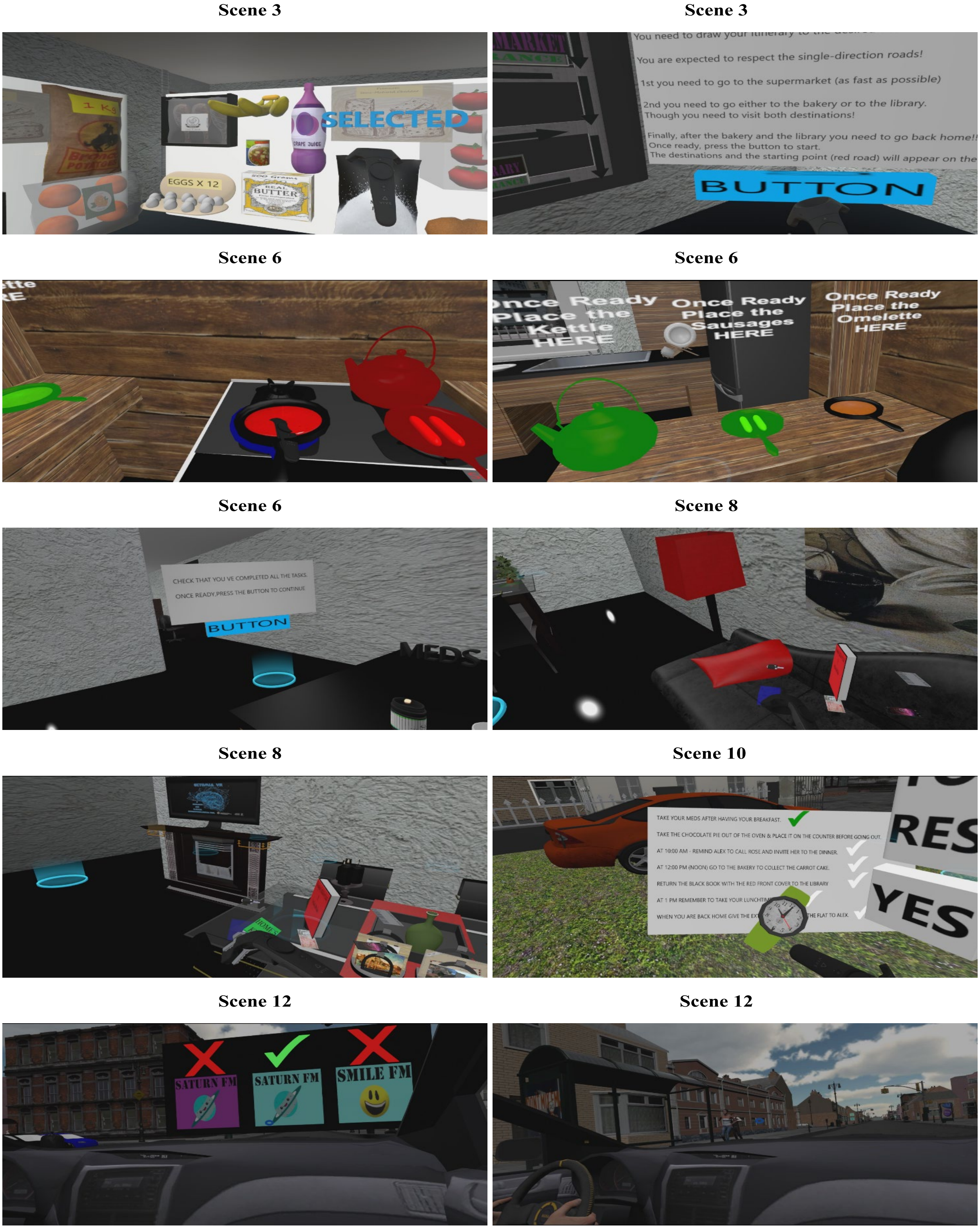
Table 3. VR-EAL tasks' scoring

Cognitive Function	Scoring
Prospective Memory	An example: At the end of a scene, the examinees should press a button to confirm that all the tasks in the scene are completed. If the examinees have already taken their medication (i.e., PM task) before pressing the final button, then the scene ends, and the examinees receive 6 points. Otherwise, the first prompt appears (i.e., "You Have to Do Something Else"). If the examinees then follow the prompt and take their medication, their receive 4 points. If the examinees press the final button again, then the second prompt appears (i.e., "You Have to Do Something After Having your Breakfast"). If the examinees follow this prompt and take their medication, they receive 2 points. If the examinees press the final button again, then the third prompt appears (i.e., "You Have to Take Your Meds"). If the examinees follow this prompt and take their medication, they then receive 1 point. If the examinees repress the final button without ever taking their medication, they get zero points, and the scene ends. Scores range from 0 to 6.
Episodic Memory	The examinee gains 2 points for each correctly chosen item, 1 point for choosing a qualitative or quantitative distractor, and 0 points for the false items. Scores range from 0 to 20.
Multitasking/ Task-Shifting	The score for each pan depends on the time that the examinee removes the pans from the stove and places them on the kitchen worktop. Equally, the score for boiling the kettle is measured in relation to the stage of the audio playback (e.g., the kettle whistles when the water is ready) that the kettle is placed on the kitchen worktop. Scores range from 0 to 16.
Planning	When the examinee selects a unit, 1 point is awarded. The ideal route to visit all three destinations is 15 units; hence, any extra or missing units are subtracted from the total possible score of 15. Up to 4 more points are awarded for the time taken to complete the task. Scores range from 0 to 19.
Visuospatial attention	The examinee receives 2 point for each target item collected (6 target items). Also, up to 4 points are awarded for the speed of detecting the items. If the examinee attempts to collect one of the distractors, it counts as an error. Up to 4 points are awarded for the accuracy of detecting items. Scores range from 0 to 20.
Visual attention	The examinee is awarded 1 point when a target poster is "spotted" and subtracted 1 point when a distractor poster is "spotted". Scores range from 0 to 16.
Auditory attention.	If the examinee presses the trigger on the right controller to detect a target sound originating on the right side (i.e., controller and sound on the same side), then s/he gets 2 points. If the examinee presses the trigger on the right controller to detect a target sound originating on the left side (i.e., controller on the opposite side), s/he gains only 1 point. If the examinee responds to a distractor sound, irrelevant of its origin or the controller used to respond, 1 point is deducted. Scores range from 0 to 32.

Table 4. VR-EAL prospective memory tasks and comparisons

Type	Description
Focal event-based tasks.	There are two focal event-based tasks, where the first task is performed in an early stage of the scenario (15-30 minutes after encoding the intention), and the second one is performed in a late stage (45-60 minutes after encoding the intention). The environmental cue in focal event-based tasks should be easily detected (i.e., focal) and of high salience (e.g., to be in the same semantic category). For example, the examinee is in front of the library facing the large sign displaying "Library", where the examinee should remember to return a book.
Non-focal event-based tasks.	Comparable to the focal event-based tasks, there are two non-focal event-based task which are performed in an early and late stage correspondingly. However, in the non-focal event-based tasks, the focality and salience of the cue should be adequately low to motivate the examinee to monitor the area for cues. For example, the examinee finishes preparing breakfast (see the cooking task below), after having breakfast the examinee should remember to take her/his medicine (see the scoring example above). The medicine is out of the field of view of the examinee, but it is in the same room (i.e., kitchen) among other kitchen objects (e.g., cups, plates, towels, microwave, and toaster). Hence, the examinee should purposefully monitor the area to detect the medicine and perform the PM task.
Time-based tasks	There are four time-based tasks, from which, three are traditional time-based tasks (e.g., call Rose at 12pm) and one is a misleading task (see below). The examinee should be checking her/his digital watch, which is attached to the left hand/controller in the virtual environment, to perform the time-based task. Two of the traditional time-based tasks, one performed in an early stage (i.e., 15-30 minutes after encoding the intention) and one in a late stage (i.e., 45-60 minutes after encoding the intention), are used for the comparison with the focal and non-focal event-based tasks.
Misleading tasks.	There are two misleading tasks. One is related to an event (i.e., event-based) and the other to a specific time (i.e., time-based). However, in the misleading tasks, the examinee should respond negatively to the prompts by a virtual human to perform a task, because there is not a task to perform. The misleading tasks serve two purposes. The first purpose of this deception is to examine whether the user is simply responding affirmatively to all PM task prompts. If the user responds affirmatively, then s/he loses points (see Table 1). The second purpose is that these false prompts are common in everyday life (e.g., a friend invites you to watch together an episode of a TV series, but you refuse because you remember that the episode will be aired tomorrow and not today). Hence, their inclusion increases the ecological validity of VR-EAL. Nevertheless, in this study, the misleading tasks are only considered for the PM total score, and not for the comparisons between PM tasks.
Early & Late tasks	The PM tasks are also divided into early and late tasks, where the early tasks are performed approximately 15-30 minutes after forming the PM intention (i.e., encoding), and the late tasks are performed approximately 45-60 minutes after forming the PM intention. This allows a comparison between the performance on early and late PM tasks (i.e., focal event-based, non-focal event-based, and time-based tasks). However, for this comparison, only one early and one late task comparison from each type is considered, while the misleading tasks are not considered.

Figure 1. VR-EA L Storyline: Scenes 3 - 12.



Scene 3

Scene 3

Scene 6

Scene 6

Scene 6

Scene 8

Scene 8

Scene 10

Scene 12

Scene 12

You are expected to respect the single-direction roads!

1st you need to go to the supermarket (as fast as possible)

2nd you need to go either to the bakery or to the library.
Though you need to visit both destinations!

Finally, after the bakery and the library you need to go back home!!
Once ready, press the button to start.

CHECK THAT YOU VE COMPLETED ALL THE TASKS.
ONCE READY,PRESS THE BUTTON TO CONTINUE

TAKE YOUR MEDS AFTER HAVING YOUR BREAKFAST. ✓
TAKE THE CHOCOLATE PIE OUT OF THE OVEN & PLACE IT ON THE COUNTER BEFORE GOING OUT.
AT 10:00 AM - REMIND ALEX TO CALL ROSE AND INVITE HER TO THE DINNER.
AT 12:00 PM (NOON) GO TO THE BAKERY TO COLLECT THE CARROT CAKE.
RETURN THE BLACK BOOK WITH THE RED FRONT COVER TO THE LIBRARY

Derived from Kourtesis et al., (2020b).

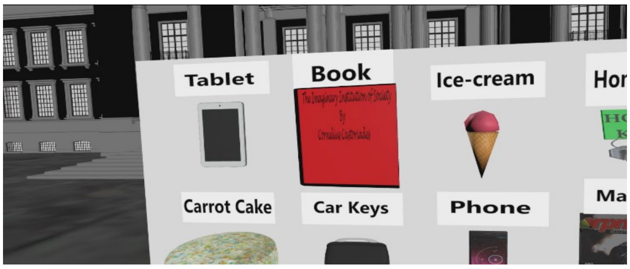
Figure 2. VR-EAL Storyline: Scenes 14 – 22

Scene 14

Scene 14

Scene 15

Scene 17



Scene 19

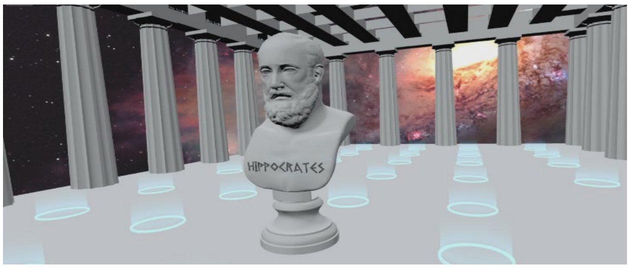
Scene 19

Scene 20

Scene 22

Scene 22

Scene 22



Derived from Kourtesis et al., (2020b).

Statistical Analyses

The Bayesian factor (BF_{10}) was used for assessing statistical inference. The BF_{10} threshold ≥ 10 was set for statistical inference in all analyses, which indicates strong evidence in favor of the H1 (Marsman & Wagenmakers, 2017; Rouder & Morey, 2012; Wetzels & Wagenmakers, 2012) and corresponds to a p-value < 0.01 (e.g., $BF_{10} = 10$) (Bland, 2015; Cox & Donnelly, 2011; Held & Ott, 2018). BF_{10} is considered substantially more parsimonious than the p-value in evaluating the evidence against the H0 (Bland, 2015; Cox & Donnelly, 2011; Held & Ott, 2018), especially when evaluating the evidence of H1 against H0 in relatively small sample sizes (Held & Ott, 2018), as in the present study. Importantly, BF_{10} allows evidence in either direction (i.e., towards H1 and H0), and its measurement of evidence is insensitive to the stopping rule, which substantially mitigates the issue of multiple comparisons and generates reliable and more generalizable results (Dienes, 2016; Marsman & Wagenmakers, 2017; Wagenmakers *et al.*, 2018).

Bayesian paired samples t-tests were conducted to explore the differences among different PM measures (i.e., time-based, focal event-based, and non-focal event-based tasks) and examine whether they were aligned or discrepant with the relevant literature. Bayesian Pearson correlational analyses were conducted to examine associations between the scores of VR-EAL, especially with the PM scores. Bayesian linear regression analyses were performed to examine the predictors of the PM scores, where the Jeffreys–Zellner–Siow (JZS) mixed g-prior was used to select the variables for the best model. JZS has the computational advantages of a g-prior in conjunction with the theoretical advantages of a Cauchy prior (Liang, Paulo, Molina, Clyde, & Berger, 2008; Rouder & Morey, 2012). Finally, post hoc statistical power analyses of the Bayesian paired samples t-tests and Bayesian linear regression analyses were performed using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007; Faul, Erdfelder, Buchner, & Lang, 2009). All Bayesian analyses were performed using JASP (Version 0.8.1.2) (JASP Team, 2018).

Results

The descriptive statistics for the performance on the VR-EAL's PM, attention, episodic memory, and executive functioning tasks are displayed in Table 5.

Table 5. Descriptive statistics for the VR-EAL scores

	N	**Mean (SD)**	**Range**
PM Total Score (max = 48)	41	35.78 (4.73)	24-46
Total Event Based (max = 24)	41	18.15 (3.26)	8-24
Focal Event Based (max = 12)	41	10.00 (1.83)	4-12
Focal Event Based – Early (max = 6)	41	5.37 (0.92)	2-6
Focal Event Based – Late (max = 6)	41	4.63 (1.26)	1-6
Non-Focal Event Based (max = 12)	41	9.27 (2.00)	5-12
Non-Focal Event Based – Early (max = 6)	41	5.22 (1.01)	2-6
Non-Focal Event Based – Late (max = 6)	41	4.05 (1.84)	0-6
Total Time Based (max = 18)	41	11.63 (3.10)	6-18
Time Based – Comparison (max = 12)	41	9.29 (1.97)	4-12
Time Based – Comparison Early (max = 6)	41	5.17 (1.34)	0-6
Time Based – Comparison Late (max = 6)	41	4.12 (1.29)	1-6
Immediate Recognition (max = 20)	41	15.51 (1.98)	10-18
Delayed Recognition (max = 20)	41	17.17 (2.42)	12-20
Visual Attention Accuracy (max = 32)	41	22.98 (3.84)	17-30
Visual Attention Speed (max = 32)	41	23.61 (3.69)	18-30
Visuospatial Attention Speed (max = 16)	41	11.90 (1.50)	8-14
Visuospatial Attention Accuracy (max = 16)	41	12.10 (1.18)	8-13
Auditory Attention (max = 32)	41	29.56 (3.66)	20-32
Planning (max = 19)	41	14.90 (1.51)	11-17
Cooking Task (max = 16)	41	9.68 (2.57)	2-13

“Time Based – Comparison”, “Time Based – Comparison Early”, and “Time Based – Comparison Late” were used for the comparisons with the corresponding “Focal Event Based” and “Non-Focal Event Based” tasks. “Early” and “Late” corresponds to the time in the scenario that this PM task was performed.

Comparisons between the different types of PM tasks

Table 6 demonstrates the VR-EAL PM scores. No significant differences were observed among the focal event-based, non-focal event-based, and time-based tasks. Despite the absence of significant differences, performance on the focal event-based tasks was higher than the non-focal event-based and time-based tasks, especially in the tasks performed during the late stage of the scenario. Performance was significantly higher on the tasks performed during the early stage of the scenario compared to the late stage of the scenario. Also, performance on the non-focal event-based task was significantly lower in the late stage compared to performance on the focal event-based and time-based tasks performed in the early stage. Equally, performance on the time-based task performed during the late stage was significantly lower than performance on the focal and non-focal event-based tasks performed in the early stage. However, late focal event-based tasks did not show a significant difference with early non-focal event-based and time-based tasks, albeit that performance on the specific focal event-based task was lower than performance on the other two tasks.

Table 6. Comparison between the VR-EAL prospective memory tasks and scores

Comparison			**p-value**	**t**	**BF_{10}**	**error%**	**SP**
Focal Event Based	>	Non-Focal Event Based	0.065	1.549	0.942	~ 3.874e -5	100 %
Focal Event Based	>	Time Based – Comparison	0.039	1.810	1.427	~ 1.093e -4	100 %
Non-Focal Event Based	>	Time Based – Comparison	0.521	0.054	0.162	~ 5.508e -6	6.32 %
Focal Event Based - Early	>	Non-Focal Event Based – Early	0.269	0.621	0.293	~ 0.005	97.25 %
Focal Event Based - Early	>	Time Based - Comparison Early	0.205	0.831	0.367	~ 2.035e -5	99.93 %
Non-Focal Event Based - Early	>	Time Based - Comparison Early	0.428	0.183	0.195	~ 5.495e -6	20.81 %
Focal Event Based - Late	>	Non-Focal Event Based –Late	0.052	1.668	1.133	~ 6.915e -5	100 %
Focal Event Based - Late	>	Time Based - Comparison Late	0.032	1.908	1.685	~ 1.372e -4	100 %
Non-Focal Event Based - Late	>	Time Based - Comparison Late	0.577	0.197	0.146	~ 3.610e -6	23.37 %
Focal Event Based - Early	>	Focal Event Based – Late	< .001	3.824	123.801 ***	~ 3.855e -5	100 %
Non-Focal Event Based - Early	>	Non-Focal Event Based – Late	< .001	3.406	41.867 **	~ 1.666e -5	100 %
Time Based - Comparison Early	>	Time Based - Comparison – Late	< .001	3.847	131.608 ***	~ 3.750e -5	100 %
Non-Focal Event Based - Early	>	Focal Event Based – Late	0.021	2.096	2.357	~ 2.001e -4	100 %
Time Based - Comparison Early	>	Focal Event Based – Late	0.028	1.966	1.865	~ 1.546e -4	100 %
Focal Event Based - Early	>	Non-Focal Event Based – Late	< .001	3.782	110.788 ***	~ 3.988e -5	100 %
Time Based - Comparison Early	>	Non-Focal Event Based – Late	0.002	3.132	21.414 *	~ 7.988e -6	100 %
Focal Event Based - Early	>	Time Based - Comparison – Late	< .001	4.999	3406.650 ***	~ 1.417e -7	100 %
Non-Focal Event Based - Early	>	Time Based - Comparison – Late	< .001	4.453	703.315 ***	~ 4.009e -6	100 %

* *$BF_{10} > 10$;* ** *$BF_{10} > 30$;* *** *$BF_{10} > 100$; SP = Statistical Power at α = .05; For all tests, hypothesis is measurement one greater than measurement two.*

Correlations between the VR-EAL scores

The only significant correlations were found between delayed recognition and the total PM score, and between visual attention accuracy and non-focal event-based score (see Table 7).

Table 7. Bayesian correlations between the VR-EAL scores

		Focal Event Based	Non-Focal Event Based	Event Based	Time Based	Total PM Score
Immediate Recognition	p-value	0.547	0.172	0.077	0.691	0.140
	r	0.097	0.217	0.279	0.064	0.234
	BF_{10}	0.232	0.478	0.879	0.210	0.556
Delayed Recognition	p-value	0.598	0.347	**0.007**	0.132	**0.004**
	r	0.085	0.151	**0.412**	0.239	**0.441**
	BF_{10}	0.222	0.298	**6.217**	0.581	**10.759***
Visual Attention Accuracy	p-value	0.211	**0.003**	0.151	0.361	0.110
	r	0.199	**0.449**	0.228	0.147	0.253
	BF_{10}	0.414	**12.695***	0.526	0.291	0.667
Visual Attention Speed	p-value	0.309	**0.006**	0.142	0.325	0.095
	r	0.163	**0.419**	0.233	0.158	0.264
	BF_{10}	0.320	**6.993**	0.551	0.311	0.746
Visuospatial Attention Accuracy	p-value	0.564	**0.050**	0.588	0.079	0.449
	r	0.093	**0.308**	0.087	0.278	0.122
	BF_{10}	0.228	**1.245**	0.224	0.864	0.257
Visuospatial Attention Speed	p-value	0.151	0.759	**0.038**	0.411	0.389
	r	0.228	0.049	**0.326**	0.132	0.138
	BF_{10}	0.526	0.204	**1.568**	0.270	0.278
Auditory Attention	p-value	**0.023**	0.429	0.091	0.102	**0.023**
	r	**0.354**	0.127	0.267	0.259	**0.354**
	BF_{10}	**2.363**	0.263	0.773	0.708	**2.339**
Planning	p-value	0.098	0.131	0.121	**0.030**	**0.011**
	r	0.262	0.240	0.246	**0.339**	**0.392**
	BF_{10}	0.728	0.587	0.623	**1.883**	**4.298**
Cooking Task	p-value	**0.046**	0.422	0.198	0.278	0.108
	r	**0.313**	0.129	0.205	0.173	0.255
	BF_{10}	**1.327**	0.266	0.434	0.343	0.679

The alternative hypothesis specifies that the correlation is positive. * *$BF_{10} > 10$;* ** *$BF_{10} > 30$;* *** *$BF_{10} > 100$*

r = Pearson's r; PM = prospective memory; In bold, there are significant and tended to be significant correlations.

Predictors of PM scores

We provided a brief description of the covariates and model selection of our Bayesian regression analyses in JASP (see Liang *et al.*, 2008; Rouder & Morey, 2012). The very first step was the selection of the covariates which form the full model. In our case, each full model included all the non-PM VR-EAL scores. JASP then formed sub-models with all the possible combinations of the covariates included in the full model, and then compared these sub-models to select the best one. Of note, one of the possible models is the null model (i.e., no covariates), which postulates that none of the covariates and their combinations substantially explained the variance of the dependent variable. For the inclusion of the covariates in the model, JASP uses the BF-inclusion, which is the ratio of the prior inclusion probability and the posterior inclusion probability. For the selection of the best model from all possible sub-models (including the full model and the null model), JASP uses the BF_M, which is the change from prior model odds to posterior model odds.

Table 8 demonstrates the predictive models for the PM scores. In all models, the variance inflation factor and tolerance were acceptable, suggesting an absence of multicollinearity issues in the models. The most frequent predictors in the five models of the PM scores (i.e., included in the 4 best models) were delayed recognition, visuospatial accuracy, multitasking/task-switching (i.e., the cooking task), and auditory attention; closely followed (i.e., included in the 3 best models) by the planning task. All models achieved good to excellent statistical power (i.e., ≈ 83 - 99%), while they explained a significant percentage of the variance on the PM scores (i.e., ≈ 26 - 47%). The best predictive model for the focal event-based score included auditory attention, multitasking/task-switching, and visual attention speed. The best model for the non-focal event-based score included visual attention accuracy, visuospatial attention accuracy, planning, and delayed recognition. The best predictive model for the event-based score included delayed recognition, visuospatial attention accuracy and speed, multitasking/task-switching, and auditory attention. Lastly, the predictive models for the time-based score and PM total score included the same predictors: delayed recognition, visuospatial attention accuracy, auditory attention, multitasking/task-switching, and planning.

Table 8. Predictive models of prospective memory scores

Target	Best Model	P(M)	P(M/Data)	BF_M	R^2	SP
Focal Event Based	Auditory Attention + Cooking Task + Visual Attention Speed	0.002	0.041	26.644	0.256	83.21%
Non-Focal Event Based	Visual Attention Accuracy + Visuospatial Attention Accuracy + Planning + Delayed Recognition	0.002	0.014	12.198	0.358	94.13%
Event Based	Delayed Recognition + Visuospatial Attention Speed + Auditory Attention + Cooking Task + Visuospatial Attention Accuracy	0.002	0.015	12.535	0.301	86.27%
Time Based	Planning + Visuospatial Attention Accuracy + Auditory Attention + Delayed Recognition + Cooking Task	0.004	0.025	11.589	0.355	91.72%
Total PM Score	Delayed Recognition + Planning + Auditory Attention + Cooking Task + Visuospatial Attention Accuracy	0.002	0.042	27.596	0.469	98.80%

P = Probability; M = Model; BF_M *= Model's Bayesian Factor; SP = Model's Statistical Power at a =.05*

Discussion

The current study investigated everyday PM functioning by implementing an immersive VR research paradigm (i.e., VR-EAL) with enhanced ecological validity. Performance on diverse PM tasks (i.e., focal event-based, non-focal event-based, and time-based tasks) was examined and the cognitive functions and processes facilitating everyday PM functioning were explored. We will now discuss our findings pertaining to everyday PM functioning in relation to the predominant theories of PM (i.e., the PAM and multiprocess theories), and the contribution of executive functions to PM functioning.

Performance on everyday PM tasks

Performance on focal event-based tasks was higher than performance on non-focal event-tasks, which is in line with the findings of previous studies making similar comparisons (e.g., Loft & Remington, 2013; McDaniel, Shelton, Breneiser, Moynan, & Balota, 2011; Mullet *et al.*, 2013; Scullin *et al.*, 2010; Zuber *et al.*, 2019). However, in our study, this difference in performance was not significant. Equally, Loft and Remington (2013) found a non-significant difference between focal and non-focal event-based tasks when participants delayed their response to the PM cue for 600ms or longer. Loft and Remington (2013) argue that the design of laboratory PM tasks motivates participant to respond as fast as possible, while a delay seems to allow participants to refocus their preparatory processes (e.g., attentional and memory) from the ongoing task (i.e., distractor) to the PM task, substantially increasing performance on non-focal event-based tasks. The performance on focal and non-focal event-based tasks was also comparable to performance on time-based tasks. Again, this contradicts findings from the PM literature where performance on time-based tasks was poorer than focal and non-focal event-based tasks (e.g., Conte & McBride, 2018; Zuber *et al.*, 2019). These discrepancies may be attributed to the delay to perform the PM tasks after the detection of the associated cue or prompt, which is potentially due to the ecological nature of the VR-EAL.

VR-EAL is an immersive and realistic simulation of a day (i.e., 8am – 3pm), where the participant has to perform several everyday tasks (i.e., both PM and non-PM tasks) according to a scenario. There are several factors that make the VR-EAL significantly closer to real life situations and PM functioning than other PM paradigms, such as the complexity and synthesis of the environment (i.e., a realistic and complex 360° environment), the quality of the stimuli (i.e., dynamic and realistic), the duration of the experiment (i.e., approximately 70 minutes), the naturalistic and ergonomic interactions within the testing environment, the response type (i.e., naturalistic) and time (i.e., in seconds), the length of the delay from encoding the PM intention to performing the PM action (i.e., 15 – 60 minutes), as well as the quantity (i.e., several) and quality (i.e., cognitively demanding) of the ongoing and PM tasks. We have previously demonstrated that participants rate the similarity of VR-EAL's tasks with real-world tasks as very high (Kourtesis *et al.*, 2020a). By adopting an ecologically valid paradigm like the VR-EAL, participants are thought to respond in a more realistic way, rather than responding almost automatically (i.e., measured in milliseconds) on laboratory tasks. For example, in VR-EAL, participants responded several seconds (or even a couple of minutes) after the presentation or detection of the PM prompt or cue. This delay suggests that, in everyday PM functioning, individuals inhibit impulsive responses related to the PM task, and/or shift their attention to the PM task, and/or check their stored information (e.g., a plan of action) to decide which response would be appropriate. However, laboratory-based PM tasks tend to adopt a single ongoing task, and a single PM cue and task (e.g., Gonneaud *et al.*, 2011; Loft & Remington, 2013; McDaniel *et al.*, 2011; Mullet *et al.*, 2013; Scullin *et al.*, 2010; Schnitzspahn *et al.*, 2013; Zuber *et al.*, 2016, 2019), which substantially reduces the requirement of executive functions on PM functioning. Hence, the observed contributions of updating, shifting, and inhibition on PM functioning in studies which have implemented such laboratory tasks may be underestimated (e.g., Gonneaud *et al.*, 2011; Schnitzspahn *et al.*, 2013; Zuber *et al.*, 2016, 2019).

Another reason for the discrepancy between the current findings and previous studies may be the implementation of Bayesian factors for significance testing instead of p-values (i.e., the null hypothesis significance testing; NHST). Reporting failures to replicate seems crucial for scientific progress (Ioannidis, 2015), while one of the main causes for failing to replicate results appears to be the false positive findings through utilizing p-values for testing significance (Ioannidis, 2005, 2014, 2015; Benjamin *et al.*, 2018). Research has shown that Bayesian statistics are substantially more parsimonious and reliable than NHST (Benjamin *et al.*, 2018; Dienes, 2016; Held & Ott, 2018; Marsman & Wagenmakers, 2017; Wagenmakers *et al.*, 2018). In our study, some of our comparisons would have been labelled “significant” (e.g., focal event-based against time-based) if we had adopted a NHST approach using a threshold of $p < .05$. The utilization of Bayesian statistics rather than NHST assured more robust and reliable results but may have contributed to our discrepant findings compared to previous studies using NHST and a threshold of $p < .05$.

The only significant differences that we found were pertaining to the duration between encoding the PM intention and performing the PM action. Performance on time-based, focal, and non-focal event-based tasks performed during the early stage (i.e., 15 – 30 minutes after encoding) was substantially higher than performance on the corresponding tasks (e.g., time-based early > time-based late) performed during the late stage (i.e., 45 – 60 minutes after encoding). This indicates that PM declines over delays of 45 minutes or longer. Our findings are concordant with the findings of previous studies that have attempted to examine the effect of delay on focal (e.g., Kliegel & Jager, 2006; Meier, Zimmermann, & Perrig, 2006; Scullin & McDaniel, 2010) and non-focal even-based tasks (e.g., Martin, Brown, & Hicks, 2011; McBride, Beckner, & Abney, 2011), as well as on time-based tasks (McBride, Coane, Drwal, & LaRose, 2013).

It has been argued that a delay may leave PM performance intact or even increase performance, if the ongoing task permits rehearsal (Hicks, Marsh, & Russell, 2000; Martin *et al.*, 2011) and does not induce a moderate to high working memory load (Kliegel & Jager, 2006). In VR-EAL, the ongoing tasks are highly engaging and demanding like real-life tasks, so they are not thought to permit rehearsal. Indeed, our findings are in line with studies where the duration of the delay (e.g., 12 hours) was more representative of everyday PM (e.g., McBride *et al.*, 2013; Scullin & McDaniel, 2010). These findings postulate the important role of episodic memory in everyday PM functioning for maintaining and retrieving the PM intention. Also, performance on late time-based and non-focal event-based tasks was significantly lower than on other early PM tasks (e.g., non-focal late < focal even-based early or time-based early). However, performance on the late focal event-based task was not significantly lower than on other early PM tasks (e.g., focal late < non-focal early or time-based early). In line with the multiprocess theory, this suggests that the role of episodic memory in focal event-based tasks, which require reflexive associative retrieval, is not as crucial as it is in non-focal and time-based tasks which require intentional retrieval (McDaniel & Einstein, 2000, 2007). Indeed, the pattern of PM forgetting seems comparable to episodic memory forgetting on non-focal event-based (Martin *et al.*, 2011; McBride *et al.*, 2011) and time-based tasks (McBride *et al.*, 2013), which indicates the critical importance of episodic memory in these PM tasks.

Cognitive functions facilitating everyday PM

The VR-EAL subtests assessing auditory attention, multitasking/task-switching, and visual attention speed significantly predicted performance on the focal event-based tasks. Auditory and visual attention may assist with the encoding of the PM intention and its automatic retrieval (McDaniel & Scullin, 2010). However, automatic retrieval is predominantly facilitated by the focality and salience of the PM cue (McDaniel & Scullin, 2010; McDaniel et al., 2015). The involvement of visual attention speed appears to be in line with the multiprocess framework, where the focality of the cue (i.e., easily detected) appears to diminish the role of attentional accuracy on PM performance, and highlights the importance of attentional speed (i.e., how fast the cue is detected). Similarly, the presence of auditory attention in the model may also indicate that the ability to detect auditory cues (e.g., the virtual character who accompanies the participant mentions the word “Library”) also assists in the retrieval of the PM intention as the visual cues do.

Also, the inclusion in the focal event-based model of a visual attention score rather than a visuospatial attention score seems to support the multiprocess framework. This framework suggests that monitoring is sometimes passive/automatic and not always intentional/strategic as the PAM theory suggests (McDaniel & Einstein, 2000; 2007; McDaniel et al., 2015). Specifically, the automatic monitoring and retrieval takes place when the PM cue is focal (i.e., the PM cue is easily detected; McDaniel & Einstein, 2000; 2007; McDaniel et al., 2015). Visuospatial attention refers to the ability to intentionally select and scan a specific area in the surrounding environment (i.e., 360°) where a specific cue may be detected, while the visual attention speed measure refers to the speed of detecting a cue within the visual field. Hence, the inclusion of visual attention speed instead of a visuospatial attention score in our predictive model for the focal event-based score might support the distinction between strategic and automatic monitoring as suggested by the multiprocess framework.

The inclusion of multitasking/task-switching in the focal event-based model highlights the role of executive functions in PM functioning. The multitasking/task-switching ability was assessed by a cooking task. Cooking tasks are composite assessments of executive functioning processes such as updating stored information, inhibiting responses discordant to the plan of action, and switching attention between tasks (Craik & Bialystok, 2006; Doherty *et al.*, 2015; Rose *et al.*, 2015), which have been found to contribute to PM functioning (Azzopardi *et al.*, 2017; Gonneaud *et al.*, 2011; Schnitzspahn *et al.*, 2013; Zuber *et al.*, 2016, 2019). Specifically, updating and inhibition have been found to facilitate focal event-based PM (Zuber *et al.*, 2016, 2019).

The predictive model for the non-focal event-based score included the scores for visual attention accuracy, visuospatial attention accuracy, planning, and delayed recognition. The inclusion of both these attentional processes emphasizes the importance of preparatory attentional processes to facilitate the strategic monitoring for non-focal PM cues (Smith, 2003; Smith *et al.*, 2007). Here both the ability to select and scan a specific area in the surrounding environment for a cue, and the ability to detect the correct cue within the visual field, play a role. The presence of visual attention accuracy and visuospatial accuracy in the non-focal model supports the distinction between automatic and strategic monitoring as suggested by the multiprocess framework.

The non-focal model also included delayed recognition. The PAM theory highlights the importance of preparatory memory processes (Smith, 2003; Smith *et al.*, 2007), while the multiprocess theory suggests that memory processes are only required for non-focal cues. Focal cues do not require memory processes because of the high salience between the cue and the PM intended action (McDaniel & Einstein, 2000, 2007; McDaniel et al., 2015). Hence, the inclusion of delayed recognition only in our non-focal model (and not in the focal model) provides further evidence for the differentiation between these two PM models (i.e., automatic retrieval with focal cues and intentional retrieval with non-focal cues).

Lastly, the non-focal model also included planning. Planning can be described as future thinking that organizes and prioritizes future actions (Morris & Ward, 2004). In everyday life, planning defines when and where an action will take place (e.g., tomorrow I will pay my overdue electricity bill after work; Morris & Ward, 2004). Planning appears to be central to Gollwitzer's term *implementation* *intention*, which is an encoding technique for future actions by associating a cue with the intended action (i.e., “if I encounter X then I will do Y”; Gollwitzer, 1999). This encoding technique has been found to significantly ameliorate PM performance (Liu & Park, 2004; McDaniel & Scullin, 2010; Milne, Orbell, & Sheeran, 2002). In focal tasks, the cue is highly associated with the intended action; however, in non-focal tasks, the cue is not strongly associated with the intended action. Thus, planning (i.e., when I see X, then I will do Y) emerges as a paramount factor for encoding the association between the cue and the PM action in non-focal event-based tasks.

The distinct models for the focal and non-focal event-based scores have highlighted the differences between the PAM and multiprocess theories in terms of monitoring and retrieval processes. In line with the multiprocess framework, the focal model postulates an automatic and cognitively cost-effective monitoring and retrieval. In line with both the PAM and multiprocess theories, the non-focal model hypothesizes a strategic, yet cognitively high cost, monitoring and intentional retrieval. Our current findings contradict the claim of the PAM theory that monitoring and retrieval processes are always, to some extent, strategic and intentional (Smith, 2003; Smith *et al.*, 2007). In contrast, our findings support the proposal of the multiprocess framework that suggests that monitoring and retrieval are sometimes intentional but sometimes automatic, depending on the focality of the PM cue (McDaniel & Einstein, 2000; 2007; McDaniel et al., 2015). Hence, as suggested by the multiprocess framework, there is an interplay between the two processes that occurs in real-life PM functioning. Our findings also pinpoint the importance of executive functions in both focal and non-focal event-based PM in encoding and maintaining the PM intention. However, they may also facilitate the interplay between intentional and automatic monitoring and retrieval. This could be seen as an extension of Gollwitzer's implementation intention; planning ability may assist with defining when shifting from cost-effective automatic monitoring to intentional and strategic monitoring will occur (e.g., the Z environmental cue was automatically detected, which provides a hint for the existence of the X cue, whereas the X cue is associated with the Y PM action). For example, on my way back home, I passively detect a huge billboard (i.e., Z cue) which indicates that I am entering the commercial area of the town (i.e., the hint for an X cue in the area) and this reminds me that I want to buy something. So, I start strategically monitoring the area until I spot the bakery sign (i.e., X cue), which reminds me that I want to buy a loaf (i.e., Y PM action).

In the model for the event-based total score which included two focal tasks, two non-focal tasks, and one misleading task, there were two predictors from the focal model (i.e., auditory attention and multitasking/task-switching), two predictors from the non-focal model (i.e., visuospatial attention accuracy and delayed recognition), and visuospatial attention speed. Hence, this model appears to integrate the focal and non-focal models. However, the visuospatial attention accuracy prevailed as a more significant predictor than visual attention accuracy, as well as visuospatial attention speed over visual attention speed. Nevertheless, this preference in favor of visuospatial attention is justified by the fact, as described above, that these scores embed information (i.e., accuracy and speed of scanning) and are related to visual attention scores. The cooking task is related to updating stored information (e.g., plan an action), inhibiting responses discordant with the plan of action, and switching attention between tasks. The cooking task was included in the event-based total score model, while the planning task was not.

Similarly, the predictive model for the time-based total score included delayed recognition, visuospatial attention accuracy, auditory attention, the cooking task, and planning. Episodic memory has been seen to play an important role in retrieving time-based intention (Einstein & McDaniel, 1996; Mackinlay, Kliegel, & Mäntylä, 2009; McFarland & Glisky, 2009); hence, the inclusion of delayed recognition as a predictor in the time-based model supports these findings. The incorporation of visuospatial attention accuracy in the time-based model is interpreted as an indication of effective time monitoring. In VR-EAL, the watch is attached to the left hand/controller of the participant which is accessible at all stages of the scenario (see scene 10 in Figure 1). In relation to the location of the hand/controller, most of the time the watch is not within the visual field of the participant (see Figures 1 and 2), while time monitoring requires the participant's effort to focus on the watch. Hence, time monitoring can be only achieved when the participant detects and focuses on the watch. Consequently, the inclusion of visuospatial attention accuracy is accordant with the findings highlighting the role of time monitoring on time-based PM (McFarland & Glisky, 2009; Mioni & Stablum, 2014; Vanneste, Baudouin, Bouazzaoui, & Taconnat, 2016).

As discussed previously, the inclusion of auditory attention in the time-based model implicates auditory attention in encoding the PM intention and detecting relevant reminding cues (e.g., the word "time" mentioned by the virtual character). Also, the time-based model included the executive functions (i.e., planning and multitasking/task-shifting). These have been found to facilitate time-based PM functioning (Azzopardi *et al.*, 2017; Gonneaud *et al.*, 2011; McFarland & Glisky, 2009; Mioni & Stablum, 2014; Vanneste *et al.*, 2016; Zuber *et al.*, 2019), and highlights the agreement of our findings with this literature. The time-based model also indicates the similarities between time-based and non-focal event-based tasks. In time-based and non-focal event-based tasks, the attentional processes are used for strategic monitoring of temporal or environmental cues respectively, while the episodic memory involvement in both models indicates an analogous intentional retrieval process. Indeed, performance on both time-based and non-focal event-based tasks is mediated by the same frontal areas (e.g., posterior frontal cortices), indicating similarities in monitoring and retrieval processes (Cona, Arcara, Tarantino, & Bisiacchi, 2012; Cona *et al.*, 2015; Gonneaud *et al.*, 2014). However, the main distinction between time-based and non-focal event-based tasks is that monitoring relates to temporal cues (i.e., time estimation and monitoring) in the time-based tasks and mainly to environmental cues in the non-focal event-based tasks, which also shows activation of diverse regions in the former (e.g., dorsolateral prefrontal cortex) and in the latter (e.g., occipital lobe; Cona *et al.*, 2012, 2015; Gonneaud *et al.*, 2014).

Finally, the predictive model for the PM total score includes the same predictors as the time-based model. The inclusion of these predictors which explain a high percentage of the variance (i.e., 47%), further supports the importance of these cognitive functions in everyday PM functioning. Our findings hence postulate that real-life PM functioning is predominantly facilitated by episodic memory (i.e., delayed recognition), visuospatial attention (i.e., visuospatial attention accuracy), auditory attention, and executive functions (i.e., multitasking/task-shifting and planning). Interestingly, these cognitive functions were also found to be the most frequent predictors in the five models of the PM scores. The delayed recognition and the visuospatial accuracy were found to be significant predictors in the same four models (i.e., non-focal, event-based, time-based, and total PM score). Likewise, the auditory attention and multitasking/task-switching were found in the same four models (i.e., focal, event-based, time-based, and total PM score). Lastly, planning was found in the three best models (i.e., non-focal, time-based, and total PM score). As discussed above, the inclusion of attentional and memory processes is explained by the PAM and multiprocess theoretical frameworks. However, these theories do not explain time-based PM, or the specific role of executive functions (i.e., planning and multitasking/task-switching) on PM functioning. Time-based PM appears proximal to non-focal event-based PM, and executive functions were seen as contributors in both encoding and maintaining PM intentions, and regulating the interplay between automatic and strategic monitoring and retrieval. The dual pathways neurocognitive model (see Figure 1 in McDaniel *et al.*, 2015), which is based on the multiprocess framework, describes the interplay between intentional and automatic monitoring and retrieval when the PM cue is non-focal or focal respectively, and seems the closest model in depicting our findings and conclusions. However, we would also suggest the integration of time-based PM (e.g., proximal to non-focal) and the role of executive functions in the dual pathways model to better illustrate the complexity of PM.

Methodological considerations and future studies

This study has some limitations. The sample size was relatively small, although Bayesian statistics allowed us to facilitate sound statistical analyses with high statistical power. Future studies should include a larger population which would allow a factorial analysis of the components of PM. Also, the participants were only young adults. A more diverse population including both younger and older adults may help to elucidate the effect of aging on PM assessed using the VR-EAL. While everyday PM functioning should be studied by implementing immersive VR simulations (e.g., VR-EAL) or real-world tasks, the VR-EAL has the limitation that it cannot measure time monitoring. A future version of the VR-EAL should integrate a mechanism that allows for the direct quantification of time monitoring. Furthermore, while ecologically valid paradigms are suitable for the investigation of real-life PM, they may be susceptible to confounding factors and fail to thoroughly examine specific processes. Laboratory tasks have the advantage of isolating confounding factors and allowing the examination of specific processes. However, laboratory PM tasks in their current form suffer from certain limitations (e.g., a two-dimensional environment, keyboard-based responses, static stimuli, and a lack of realism). A solution would be the modernization of laboratory tasks using immersive VR technology, which has the advantage of reducing the divergence from real-life conditions, and enabling a meticulous examination of the PM components, cue attributes, and the role of executive functions.

Conclusions

In summary, our findings postulate that the performance differences between focal, non-focal, and time-based are not significant in everyday PM. In contrast, the length of delay between encoding and retrieving the PM intention appears to play a central role in everyday PM functioning. Our results pertaining to non-focal and focal event-based PM support both the PAM and multiprocess frameworks, respectively. However, our findings inferring a dynamic interplay between automatic and intentional monitoring and retrieval processes, agree with the inclusive approach of multiprocess framework. We also found that real-life PM functioning is predominantly facilitated by episodic memory, visuospatial attention, auditory attention, and executive functions. Future research should attempt to further investigate everyday PM functioning, especially time-based PM, and the role of executive functions in PM.

Chapter 6 showed that the real-world prospective memory is facilitated by episodic memory, attentional processes, and executive functions such as planning and multitasking. Also, Chapter 6 demonstrated that the everyday prospective memory functioning incorporates an interplay between automatic and intentional monitoring and retrieval processes, which aligns with the multiprocess framework. However, the differences of the performance on diverse types of prospective memory tasks, which were observed in laboratory paradigms, they were not replicated in an ecologically valid paradigm like the VR-EAL. Nonetheless, the length of delay between encoding and retrieval has an impact on the prospective memory functioning. Overall, the Chapter 6 offered evidence that the VR-EAL may also contribute to the understanding of real-world prospective memory functioning. The following chapter will offer an all-inclusive discussion of the findings of this thesis pertinent to the utility of the VR-EAL.

Chapter 7: General Discussion

This final chapter provides a summary of the findings and a general discussion of the results. Limitations and future directions related to this work are also discussed.

7.1. Aims of this thesis

This thesis adopted a multidisciplinary approach (i.e., computer science and psychology) to explore the potency of immersive virtual reality (VR) as a research and clinical tool in cognitive neuroscience and neuropsychology, as well as to address the issue of ecological validity in neuropsychological testing, especially regarding the assessment of cognitive functions which are central to everyday functioning. Finally, the technical and methodological pitfalls associated with the implementation of immersive VR in cognitive neuroscience and neuropsychology were also examined. In Chapter 2, a technological systematic literature review of the reasons for adverse VR induced symptoms and effects (VRISE) was conducted. Also, suggestions and technological knowledge for the implementation of VR head-mounted displays (HMD) in cognitive neuroscience were provided. A meta-analysis of 44 neuroscientific and neuropsychological studies involving VR HMD systems was performed. Another aim was to devise a brief tool to appraise and report both the quality of software features and VRISE intensity quantitatively, due to the absence of such a tool. Chapter 3 described and discussed the development and validation of the Virtual Reality Neuroscience Questionnaire (VRNQ), a tool to assess the quality of VR software in terms of user experience, game mechanics, in-game assistance, and VRISE. In the same chapter, suggestions pertaining to the maximum duration of VR sessions were offered. In Chapter 4, guidelines were proposed for the various stages in the development of the Virtual Reality Everyday Assessment Lab (VR-EAL), the first immersive VR neuropsychological battery, programmed using Unity game development software. In Chapter 5, the convergent, construct, and ecological validity of VR-EAL as an assessment of prospective memory, episodic memory, visual attention, visuospatial attention, auditory attention, and executive functions were examined. Finally, in Chapter 6, using VR-EAL, the predominant theories of prospective memory, the preparatory attentional and memory processes (PAM) and the multiprocess frameworks, were examined by comparing performance on diverse prospective memory tasks (i.e., focal and non-focal event-based, and time-based tasks) and identifying the cognitive functions which predict everyday prospective memory functioning. The findings of these aforementioned studies have already been discussed individually in each chapter. However, in this final chapter, the results of these studies will be discussed using an all-inclusive approach.

7.2. Summary of the results

Table 1 provides a summary of the aims and the findings of each study of this thesis.

7.2.1. Table 1. The main findings of each study

Chapter	Aims	Findings
Chapter 2	Identify the technical reasons for VRISE and examine the effect of these technical factors in neuroscientific and neuropsychological studies.	• The review indicated features pertinent to display, sound, motion tracking, navigation mode, ergonomic interactions, user experience, and computer hardware that should be considered by researchers. • The meta-analysis of the VR studies demonstrated that new generation HMDs induce significantly less VRISE and marginally fewer dropouts. Importantly, the commercial versions of the new generation HMDs with ergonomic interactions had zero incidents of adverse symptomatology and dropouts. HMDs equivalent to or greater than the commercial versions of contemporary HMDs, accompanied with ergonomic interactions, are suitable for implementation in cognitive neuroscience.
Chapter 3	Develop and validate the VRNQ and explore the maximum duration of VR sessions.	• VRNQ is a valid tool for assessing VR software in terms of self-reported user experience, game mechanics, in-game assistance, and VRISE intensity; it has good convergent, discriminant, and construct validity. • The maximum duration of VR sessions should be between 55 and 70 minutes when the VR software meets or exceeds the parsimonious cut-offs of the VRNQ, and the users are familiarized with the VR system.

Chapter 4	Provide guidelines for the development of VR software in cognitive neuroscience and neuropsychology, by describing the development of VR-EAL.	• The Unity game engine, in conjunction with compatible software incorporating assets and software development kits, assist cognitive scientists in overcoming challenges pertinent to VRISE and the quality of the VR software. • Better in-game assistance, game mechanics, and graphics substantially increase the quality of the user experience and almost eradicate VRISE. • It is feasible to develop effective VR research and clinical software without the presence of VRISE during a 60-min VR session.
Chapter 5	Validate and compare VR-EAL against a paper-and-pencil ecologically valid neuropsychological battery.	• VR-EAL scores were significantly correlated with their equivalent scores on the paper-and-pencil tests. • The participants' self-reports indicated that the VR-EAL tasks were considered significantly more ecologically valid and pleasant to perform than the paper-and-pencil neuropsychological battery. Also, the VR-EAL battery had a shorter administration time. • The VR-EAL is a suitable neuropsychological assessment of everyday cognitive functions with enhanced ecological validity, providing a highly pleasant testing experience, and not inducing cybersickness.

Chapter 6	Examine the main theoretical frameworks of focal and non-focal event-based, and time-based prospective memory using an ecological valid research paradigm, as well as to identify the cognitive functions which predict everyday prospective memory functioning.	• The type of prospective memory task does not seem to influence everyday prospective memory functioning. • The length of the delay between encoding and retrieving the prospective memory intention appears to play a central role in everyday prospective memory functioning. • The PAM and multiprocess frameworks were found to depict prospective memory functioning in non-focal and focal event-based tasks respectively. • Everyday prospective memory functioning is mediated by the dynamic interplay between automatic and strategic monitoring and retrieval processes. • The integration of time-based prospective memory and executive functions in the dual pathways model is suggested. • Everyday prospective memory functioning is predominantly facilitated by episodic memory, visuospatial attention, auditory attention, and executive functions.

As Table 1 illustrates, the implementation of immersive VR in cognitive neuroscience and neuropsychology may be efficient and advantageous. Specifically, it is feasible to avoid or substantially alleviate adverse VRISE and provide a neuropsychological assessment like VR-EAL with enhanced ecological validity and a shorter administration time. Also, the VR-EAL was rated as a highly pleasant testing experience and able to contribute to the understanding of everyday cognition.

In Chapter 2, the meta-analysis showed that the technical reasons for VRISE pertain to a number of factors: the type of display (Kim, Choe, Hwang, & Kwag, 2017); the quality and spatialization of sound (Vorländer & Shinn-Cunningham, 2014); the accuracy and speed of motion tracking (Plouzeau, Paillot, Chardonnet, & Merienne, 2015); the navigation method within the virtual environment (Porcino, Clua, Trevisan, Vasconcelos, & Valente, 2017); the type of interactions within the virtual environment (Figueiredo, Rodrigues, Teixeira, & Techrieb, 2018); the quality of in-game instructions, prompts, and tutorials (Jerald, LaViola Jr, & Marks, 2017); and the potency of the hardware (Anthes, García-Hernández, Wiedemann, & Kranzlmüller, 2016). New generation HMDs were found to induce significantly less VRISE and marginally fewer dropouts than obsolete HMDs. Notably, there were no incidents of adverse symptomatology and dropouts in studies that have used a contemporary HMD in conjunction with ergonomic interactions within the virtual environment. However, this meta-analysis also indicated that technological competency in VR is inadequate among many neuroscientists, and that researchers did not quantitatively report the quality of the VR software or the intensity of VRISE (Kourtesis, Collina, Doumas, &

MacPherson, 2019). The latter was attributed to the absence of a tool that would quantify the quality of the software and the intensity of VRISE.

In Chapter 3, the VRNQ was found to be a valid tool for assessing VR software in terms of user experience, game mechanics, in-game assistance, and VRISE intensity. The VRNQ was implemented to assess commercial VR software which incorporate the aforementioned technical details. The findings postulated that deeper immersion, better quality of graphics and sound, and more helpful in-game instructions and prompts were found to substantially reduce VRISE intensity. Hence, these findings are in agreement with the existing literature on the importance of these technical features (e.g., de Franca & Soares, 2017; Palmisano, Mursic, & Kim, 2017). Also, the overall quality of the VR software substantially modulates the maximum duration of VR sessions, while gaming experience, age, and education of the participants do not. Research involving immersive VR software should meet or exceed the parsimonious cut-offs of the VRNQ, and participants should be familiarized with the VR HMD system prior to being immersed. Meeting these criteria facilitates a maximum duration of VR session of approximately 55 to 70 minutes. However, the development of VR software is predominantly dependent on third parties (e.g., freelancers or companies) with programming and software development skills (Slater, 2018). One solution that might promote the adoption of immersive VR as a research and clinical tool might be the in-house development of VR software by computer science literate cognitive neuroscientists or research software engineers.

Chapter 4 went on to demonstrate that it is feasible to develop VR software in-house that does not result in VRISE if the cognitive neuroscientist is computer science literate. This is done using the Unity game engine, together with other software that provide assets and facilitate VR software development. The comparison amongst the versions of VR-EAL (i.e., alpha, beta, and final) postulated that better in-game assistance, game mechanics, and graphics substantially increased the quality of the user experience and almost eradicated VRISE. The final version of VR-EAL achieved high scores in every sub-score of the VRNQ and exceeded its parsimonious cut-offs. Hence, the VR-EAL, which incorporates a scenario of approximately 60 minutes, appears to be an immersive VR neuropsychological battery of everyday cognition which does not induce VRISE. As discussed in Chapter 4, the scoring criteria and design of the cognitive tasks in the VR-EAL were based on existing paper-and-pencil or non-immersive VR tests that were considered ecological valid. However, the validity of VR-EAL was not assessed against those established ecologically valid tests; this was the purpose of Chapter 5.

In Chapter 5, VR-EAL scores were significantly correlated with their equivalent scores on the ecologically valid paper-and-pencil tests, which support the convergent, construct, and ecological validity of the VR-EAL. The participants' reports indicated that the VR-EAL tasks were considered significantly more similar to real-world tasks and more pleasant than the paper-and-pencil neuropsychological battery. The VR-EAL battery also had a shorter administration time. The VR-EAL appears to be an effective neuropsychological tool for the assessment of everyday cognitive functions, which has enhanced ecological validity, a highly pleasant testing experience, and does not induce cybersickness symptomatology. Ecological validity is essential for assessing everyday cognitive functioning (Franzen & Wilhelm, 1996; Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). Notably, VR-EAL assesses prospective memory; the importance of being able to assess prospective memory has been highlighted in studies on cognitive aging (Kidder, Park, Hertzog, & Morrell, 1997), mild cognitive impairment (Schmitter-Edgecombe, Woo, & Greeley, 2009), brain injuries (Groot, Wilson, Evans, & Watson, 2002; Shallice & Burgess, 1991), human immunodeficiency viruses (HIV; Woods *et al.*, 2008), schizophrenia (Twamley *et al.*, 2008), and Parkinson's disease (Pirogovsky, Woods, Filoteo, & Gilbert, 2012).

Chapter 6 showed that immersive VR methods are capable of contributing to our understanding of prospective memory in the real-world. The VR-EAL was designed in line with the theoretical frameworks pertaining to prospective memory and the methodological guidelines for examining prospective memory by McDaniel, Umanath, Einstein, and Waldum (2015). Using the VR-EAL, the prospective memory task type (i.e., focal event-based, non-focal event-based, and time-based) does not seem to differ in terms of participants' performance, as has been observed in previous lab-based experiments (e.g., Conte & McBride, 2018; McDaniel, Shelton, Breneiser, Moynan, Balota, 2011; Mullet *et al.*, 2013; Scullin, McDaniel, Shelton, & Lee, 2010; Zuber, Kliegel, & Ihle, 2016; Zuber, Mahy, & Kliegel, 2019). Instead, the length of the delay between encoding and retrieving the prospective memory intention appears to play a central role in everyday prospective memory functioning, again this aligns with the existing literature (Kliegel & Jager, 2006; Martin, Brown, & Hicks, 2011; McBride, Beckner, & Abney, 2011; McBride, Coane, Drwal, & LaRose, 2013; Meier, Zimmermann, & Perrig, 2006; Scullin & McDaniel, 2010). The intentional and automatic monitoring and retrieval processes were found to explain prospective memory functioning in non-focal and focal event-based tasks respectively, which aligns with the PAM (Smith, 2003; Smith, Hunt, McVay, & McConnell, 2007) and multiprocess (McDaniel & Einstein, 2000; 2007; McDaniel *et al.*, 2015) theories. In line with the dual pathways model, a dynamic interplay between automatic and intentional monitoring and retrieval processes was observed (McDaniel *et al.*, 2015). However, my findings suggested that the dual pathways model should also integrate time-based prospective memory proximal to non-focal event-based prospective memory and the role of executive functions to comprehensively depict the complexity of prospective memory functioning. Concordant with the relevant literature, our findings postulated that everyday prospective memory functioning is predominantly facilitated by episodic memory (Einstein & McDaniel, 1996; Mackinlay, Kliegel, & Mäntylä, 2009; McFarland & Glisky, 2009), visuospatial attention (Smith, 2003; Smith *et al.*, 2007), auditory attention (McDaniel & Scullin, 2010), and executive functions (Azzopardi, Auffray, & Kermarrec, 2017; Gonneaud *et al.*, 2011; Schnitzspahn, Stahl, Zeintl, Kaller, & Kliegel, 2013; Zuber *et al.*, 2016; Zuber *et al.*, 2019).

In summary, the implementation of immersive VR software such as the VR-EAL appears to be valuable in cognitive neuroscience and neuropsychology. However, competence in immersive VR technology is required to avoid the pitfalls of VRISE and provide a pleasant and ecological valid assessment of everyday cognitive functions.

7.3. VR-EAL as an ecological valid neuropsychological assessment.

Bauer and collaborators (2012) published the official joint position of the American Academy of Clinical Neuropsychology (AACN) and the National Academy of Neuropsychology (NAN) which discusses 8 key issues regarding the development, dissemination, and implementation of Computerized Neuropsychological Assessment Devices (CNADs) for research and clinical purposes. The CNADs encompass any new computer-based neuropsychological assessments or computerised versions of already established paper-and-pencil tests (e.g., Wisconsin Card Sorting Test; Sahakian & Owen, 1992) or web-based tests. The CNAD could be a standalone device (i.e., hardware and software) or software (i.e., either installed locally or on the internet) that can be run on a device with processing power such as personal computers, laptops, tablets, or smartphones (Bauer, Iverson, Cernich, Binder, Ruff, & Naugle, 2012). The VR-EAL as an immersive VR software and neuropsychological assessment would be categorised as a CNAD. Hence, VR-EAL should meet the criteria of AACN and NAN to be effectively implemented for clinical or research purposes.

The AACN and NAN recognise the potential advantages of CNADs such as testing larger number of individuals quickly (e.g., parallel administration); immediately available tests; enhanced accuracy and precision (e.g., reaction time measurements); shorter administration time and reduced costs (e.g., for test administration and scoring); adaptable in different languages; exporting the data automatically (e.g., for research purposes); increased accessibility (e.g., remotely); and the integration of algorithms for making decisions on issues such as the identification of an impairment or a statistically reliable change (Bauer *et al.*, 2012). In my PhD thesis, the VR-EAL has already shown that it achieves several of these benefits. The VR-EAL is immediately available after its installation on a personal computer and automatically produces accurate performance scores that are exported into a .txt file (Kourtesis, Korre, Collina, Doumas, & MacPherson, 2020b). Consequently, the VR-EAL has also no costs for administration and scoring, and it requires a substantially shorter administration time as compared to the equivalent paper-and-pencil batteries (Kourtesis, Collina, Doumas, & MacPherson, 2020a).

However, the VR-EAL currently does not incorporate a predictive algorithm for identifying cognitive impairment, since the VR-EAL was not administered to any clinical population. Thus, the predictive validity of VR-EAL has not been established. This is considered one of the future directions, which will also allow the formulation of a predictive algorithm for identifying impairment in the targeted population. Also, the procedure for adapting VR-EAL into a different language and culture is more complex than the adaptation of a paper-and-pencil test, since, in the case of the VR-EAL, this procedure requires programming and software development skills, which will necessitate more time. Lastly, the VR-EAL may be accessible remotely, yet the unsupervised (i.e., supervision by a trained clinician or a researcher) administration of the VR-EAL is not recommended, while the installation requires hardware (i.e., immersive VR HMD, controllers, motion tracking devices, and a VR-ready personal computer) which may be unaffordable for an individual to purchase.

Nevertheless, as mentioned above, AACN and NAN specified eight issues that should be addressed to benefit from the aforementioned advantages of CNADs (Bauer *et al.*, 2012). These issues are pertaining to: (1) the safety and effectivity of the CNAD; (2) the identity of the end-user (i.e., the operator of the CNAD); (3) the technical hardware and software features of the CNAD; (4) privacy and data security; (5) the psychometric properties of the CNAD; (6) examinee issues (e.g., cultural, experiential, and disability issues); (7) the use of reporting services; and (8) the reliability of the responses and results of the CNADs (i.e., the performance on CNADs; Bauer *et al.*, 2012). Therefore, the utility of the VR-EAL should be discussed in relation with the guidelines for CNADs by AACN and NAN. The aim of this discussion is to highlight how the VR-EAL already satisfies these criteria, as well as to identify the shortcomings of the VR-EAL and define the necessary future directions for improving VR-EAL's utility as a research and clinical tool.

7.3.1. End-user, privacy, and reliability issues (points 2, 4, 7, and 8)

A critical issue is the targeted end-user of the CNAD (i.e., the person who operates the CNAD). As defined by the American Psychological Association (APA), researchers and clinicians "*do not promote the use of psychological assessment techniques by unqualified persons, except when such use is conducted for training purposes with appropriate supervision*" (APA, 2010, Ethical Standard 9.07, Assessment by Unqualified Persons). CNADs can be implemented by other professionals who do not have a background in psychometrics or neuropsychology but the results should be integrated and interpreted by a competent professional such as a cognitive neuroscientist or neuropsychologist (Bauer *et al.*, 2012). Specifically, in Chapter 4 where I discuss the development of VR-EAL, it is clearly stated that the VR-EAL should be administered by a clinician or researcher who has competency in both neuropsychological assessment and immersive VR technologies (Kourtesis *et al.*, 2020b). Therefore, the definition that the end-user of VR-EAL should be a trained professional hence aligns with the ethical principles of the APA (APA, 2010, Ethical Standard 9.07, Assessment by Unqualified Persons).

Furthermore, regarding privacy and data security, test scoring and interpretation, and record keeping, the principal concern of AACN and NAN pertains to whether the end-user would be trained to follow the respective APA guidelines and ethical standards (Bauer *et al.*, 2012). The cognitive neuropsychologist or neuroscientist administering the VR-EAL should abide with the record keeping guidelines (e.g., data should be stored and encrypted locally) of the APA (APA, 2007, Record Keeping Guidelines). The VR-EAL offers a .txt file to the end-user, where all the recorded data (i.e., response times, duration of each task, quantification of various types of errors, and cognitive performance scores) are displayed (Kourtesis *et al.*, 2020b). This .txt file and the containing data (e.g., if they have been transferred to an excel file) should be stored locally and encrypted, which is a common practice among researchers and clinicians, (APA, 2007, Record Keeping Guidelines). Moreover, since the end-user of VR-EAL should be an individual trained in psychometrics and neuropsychology, the end-user should be capable of integrating and interpreting the data amassed by VR-EAL, which also agrees with the APA ethical standards for test scoring and interpretation (APA, 2010, Ethical Standard 9.09, Test Scoring and Interpretation Services). Therefore, the guideline that every VR-EAL end-user should be a cognitive neuropsychologist or neuroscientist meets points 2 (i.e., end-user issues), 4 (i.e., privacy and data security issues), and 7 (i.e., scoring and data recording issues) of the guidelines of AACN and NAN for the appropriate implementation of CNADs.

Furthermore, examinee cooperation and sufficient motivation are crucial for obtaining reliable neuropsychological test scores (AACN, 2007; Bauer *et al.*, 2012; Heilbronner *et al.*, 2009). Specifically, participants' efforts have been found to substantially affect performance on neuropsychological tests; indeed, in some studies, participants' effort was found to have a greater impact on their cognitive performance than the pathophysiological condition (Constantinou, Bauer, Ashendorf, Fisher, & McCaffrey, 2005; Stevens, Friedel, Mehen, & Merten, 2008; West, Curtis, Greve, & Bianchini, 2011). However, when the end-user of the CNAD is a trained clinician or researcher, they are capable of identifying behavioural signs (e.g., slow movements when there is not any motor disability) that there is reduced effort by the participant through behavioural observation (Bauer *et al.*, 2012; Heilbronner *et al.*, 2009). Nevertheless, the suspicion for poor effort on cognitive tests should be further explored and confirmed (e.g., using an effort test; Bauer *et al.*, 2012; Heilbronner *et al.*, 2009). Consequently, the suggestion that the end-user of VR-EAL should be a trained clinician or researcher assists with the detection and confirmation of poor effort on the VR-EAL's tasks by the participant. In addition, the VR-EAL, as an immersive VR software which has game-like features (e.g., a user-centred interface) and simulates everyday tasks within a realistic scenario, appears to engage and motivate the examinees (Kourtesis *et al.*, 2020a, 2020b). Notably, the two different samples of participants in the studies described in Chapters 4 and 5 rated the VR-EAL as a highly pleasant testing experience (Kourtesis *et al.*, 2020a, 2020b). Motivating the participant to perform the tasks is important for acquiring reliable data, while it also assists with identifying behavioural signs of poor effort (Heilbronner *et al.*, 2009). Thus, the motivating nature of VR-EAL (i.e., a highly pleasant testing experience with an engaging scenario) may also assist with the avoidance or the detection of potential issues pertaining to the examinee's effort.

7.3.2. Technical features, safety, and effectivity issues (points 1 and 3)

The AACN and NAN underline that a CNAD should meet the safety criteria of the Federal Food, Drug & Cosmetic Act (FD&C; Bauer *et al.*, 2012). Section 201(h) of the FD&C (21 U.S.C. 301) defines a "*medical device*" as "*an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including a component part, or accessory which is . . . intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals . . . .*". Hence, a CNAD as a medical device should also comply with the safety criteria of FD&C (i.e., to not cause any harm to the examinees; Bauer *et al.*, 2012). Any inconvenience or adverse effects may be attributed to the hardware and software features of the CNAD (Cernich, Brennana, Barker, & Bleiberg, 2007; Bauer *et al.*, 2012). Likewise, the hardware and software features of a CNAD may compromise the effectivity of a CNAD and the reliability of the acquired neuropsychological and/or physiological data (e.g., Cernich *et al.*, 2007; Bauer *et al.*, 2012). Nonetheless, Parsons, McMahan, and Kane (2018) argued that contemporary hardware (e.g., personal computers with dual processors) have the computing power to sustain the parallel operation of several software, while software are now developed to exploit and effectively use this computing power. These recent technological advancements pertaining to hardware and software, allow the parallel acquisition of accurate and reliable data such as reaction times, errors, neuroimaging data, and physiological data (Parsons, McMahan, & Kane, 2018). Regarding VR-EAL, the principal problem is the presence of adverse VRISE, which compromise the safety of the participants and the reliability of the acquired data. Intense VRISE have been found to compromise overall cognitive performance (i.e., neuropsychological data;

Mittelstaedt, Wacker, & Stelling, 2019; Nalivaiko, Davis, Blackmore, Vakulin, & Nesbitt, 2015; Nesbitt, Davis, Blackmore, & Nalivaiko, 2017) and increase electrical activity and connectivity of frontotemporal and occipital lobes (i.e., neuroimaging data; Arafat, Ferdous, & Quarles, 2018; Gavgani *et al.*, 2018; Toschi *et al.*, 2017). The main cause of VRISE is the implementation of immersive VR hardware (e.g., HMDs and personal computers) of inadequate quality (e.g., low resolution or processing power) and/or software that does not have certain features (e.g., ergonomic navigation and interaction system; Kourtesis *et al*., 2019).

The meta-analysis in Chapter 2 confirmed the importance of the hardware characteristics for removing VRISE, where studies which utilized contemporary HMDs had substantially less incidents of VRISE and dropouts. Studies that used an HTC Vive HMD (Kim *et al.*, 2017) with two lighthouse stations for motion tracking (Plouzeau *et al.*, 2015) and two HTC Vive wands with six degrees of freedom (6DoF) for navigation and interactions within the virtual environment (Figueiredo *et al.*, 2018) have reduced or eradicated VRISE. In line with this, Chapters 3, 4, and 5 used hardware that was in line with the hardware-related suggestions. There were no dropouts and the presence and intensity of VRISE was minimal to none, which further confirms the importance of the suggested hardware characteristics. The acquisition of an appropriate HMD by labs or clinics is financially feasible since commercial desktop-based (e.g., HTC Vive) and standalone (e.g., Oculus Quest) HMDs can be purchased for a relatively low price (e.g., £300 - £500; Kourtesis *et al.*, 2019b). As a result, recent immersive VR studies have implemented HMDs which meet the minimum hardware characteristics (e.g., Banakou, Kishore, & Slater, 2018; Detez et al., 2019; George, Demmler, & Hussmann, 2018; Mottelson & Hornnaek, 2017;

Parsons & McMahan, 2017). Also, the VR-EAL is compatible only with these recent HMDs. Therefore, the VR-EAL appears to meet the hardware criteria of AACN and NAN, which ensure the safety of the examinees and the reliability of the acquired data.

However, beyond the hardware characteristics, the quality of the software is equally important in avoiding or alleviating VRISE incidence and intensity. Using an appropriate HMD and hardware while the software does not have the required characteristics may still result in intense VRISE and dropouts (e.g., Detez *et al.*, 2019). The navigation within the virtual environment should be facilitated by teleportation or physical movement or a combination of both (Porcino *et al.*, 2017), and the interactions with the virtual environment should be ergonomic and naturalistic (Figueiredo *et al.*, 2018). Furthermore, the in-game instructions, prompts, and tutorials should provide the user with adequate and salient information regarding the storyline, controls, and orientation (Jerald *et al.*, 2017). Lastly, the audio and ambient sounds within the virtual environment should be spatialized and of high quality (Vorländer & Shinn-Cunningham, 2014). On the basis of these recommendations, the VR-EAL adopted a navigation method combining teleportation and physical movement (Porcino *et al.*, 2017), ergonomic and naturalistic interactions (Figueiredo *et al.*, 2018), spatialized and high definition audio (Vorländer & Shinn-Cunningham, 2014), and several informative in-game instructions, prompts, and tutorials (Jerald *et al.*, 2017). In Chapter 3, the VRISE intensity was minimal and only related to fatigue. In Chapters 4 (i.e., final version) and 5, the implementation of VR-EAL showed no dropouts and the VRISE incidence and intensity was negligible and again solely related to fatigue. These replicating results in Chapters 3, 4, and 5 confirm the significance of these software features in avoiding or alleviating the incidence and intensity of VRISE, as well as demonstrating that the VR-EAL incorporates these software features and does not induce significant VRISE, which comply with the software criteria of AACN and NAN.

However, the VR software features are not only crucial for the avoidance or alleviation of VRISE, but also for the efficiency of the VR software. The ultimate purpose of VR is for the immersion to be adequately deep to deceive the brain into believing that the virtual world is the real world. The depth of immersion depends on the strength of three perceptual illusions: the placement, plausibility, and embodiment illusions (Maister, Slater, Sanchez-Vives, & Tsakiris, 2015; Pan & Hamilton, 2018; Slater, 2009; Slater, Spanlang, & Corominas, 2010). The placement illusion is the deception that the virtual environment is a real one; hence, it depends on the proximity of the appearance of the virtual environment to an equivalent real environment (Slater, 2009; Slater *et al.*, 2010). The plausibility illusion is the deception that the virtual environment reacts to the laws of physics and the actions of the participant, thus, it depends on the proximity of the virtual environment's behaviour and senses to real life (Slater, 2009; Slater *et al.*, 2010). The embodiment illusion is the deception that the virtual body of the participant is her/his own body; hence, it depends on the proximity of the virtual body' appearance and behaviour (i.e., synchronized with the movements in the physical environment) to the participant's real body and movement (Maister *et al.*, 2015; Pan & Hamilton, 2018).

Beyond the level of immersion, the three illusions (i.e., placement, plausibility, and embodiment) are also important for the ecological validity of the immersive VR CNAD. The three illusions ensure that the individual will perform the tasks as s/he would perform them in real life (Maister *et al.*, 2015; Pan & Hamilton, 2018; Slater, 2009; Slater *et al.*, 2010). The VR-EAL has substantially strong placement and plausibility illusions, and a moderate embodiment illusion. As reported by the participants, these illusions resulted in deep immersion levels in Chapters 4 and 5. Chapter 4 provided an explicit description of how the VR-EAL tasks were designed to resemble everyday life tasks (e.g., cooking, shopping, and finding items in the living room). In Chapter 5, the participants reported that the VR-EAL tasks are very similar to the corresponding tasks that they perform in everyday life. Also, the performance of the participants on the VR-EAL tasks was significantly correlated with their performance on ecological valid paper-and-pencil tasks. Hence, the findings of Chapters 4 and 5 postulated that the three illusions are crucial to ecological validity, as it has been supported by the relevant literature (i.e., Maister *et al.*, 2015; Pan & Hamilton, 2018; Slater, 2009; Slater *et al.*, 2010). Also, in line with the criteria of the Federal Food, Drug & Cosmetic Act, the software features of VR-EAL enabled the VR-EAL to efficiently achieve its purpose to deliver an ecological valid assessment of these everyday cognitive functions.

7.3.3. Psychometric properties issues (point 5)

An important issue highlighted by the AACN and NAN is that, similar to traditional psychometric tests, the CNADs abide to the same standards and conventions of psychometric test development, such as providing evidence regarding their reliability, validity, and utility (Bauer *et al.*, 2012). The information pertaining to the psychometric properties of the CNAD, which support the claimed purpose or application of the test, should be provided to potential end-users of the CNAD (Bauer *et al.*, 2012). Notably, the APA ethical standards (APA, 2010) state that "*Psychologists who develop tests and other assessment techniques use appropriate psychometric procedures and current scientific or professional knowledge for test design, standardization, validation, reduction or elimination of bias, and recommendations for use*" (Standard 9.05). Hence, all cognitive tests, either traditional or CNAD, must meet minimum psychometric standards for reliability and validity. The validity of a test examines different psychometric properties of the test such as the content validity (i.e., the test measures the cognitive domain that is supposed to measure; e.g., episodic memory), construct validity (i.e., the test measures the cognitive function(s) that is supposed to measure), and criterion-related validity (e.g., diagnostic validity, the test efficiently detects a cognitive disorder such as Alzheimer's disease; Nunnally & Bernstein, 1994). Similarly, the aspects that are examined for the reliability of a test are the internal consistency (i.e., the consistency across all the items of the test), rest-retest (i.e., consistency over time), alternate forms (i.e., consistency across all forms/versions of the test), and inter-rater reliability (i.e., consistency of the scores across diverse examiners; Nunnally & Bernstein, 1994). Importantly, as APA (2010) Ethical Standard 9.02 (Use of Assessments), Section (b) states, "*Psychologists use assessment instruments whose validity and reliability have been established for use with members of the population tested. When such validity or reliability has not been established, psychologists describe the strengths and limitations of test results and interpretation.*"

In VR-EAL, the principal aim was to develop an immersive VR neuropsychological battery with enhanced ecological validity for the assessment of cognitive functions central in everyday functioning. Hence, VR-EAL had to be consistent with the available ecological valid assessments of these everyday cognitive functions. For the development of VR-EAL, the procedures and scoring systems of established ecologically valid paper-and-pencil tests such as the Test of Everyday Attention (Robertson, Ward, Ridgeway, and Nimmo-Smith, 1996), the Rivermead Behavioral Memory Test – III (Wilson, Cockburn, & Baddeley, 2008), the Behavioral Assessment of the Dysexecutive Syndrome (Wilson, Evans, Emslie, Alderman, & Burgess, 1998), and the Cambridge Prospective Memory Test (Wilson *et al.*, 2005) were meticulously studied. However, the fact that the development of VR-EAL was based on the procedures and scoring of established ecological valid tests does not ensure that the VR-EAL will have equivalent psychometric properties. As the AACN and NAN suggest, even a computerised version of an established paper-and-pencil test should be treated as a new test, for which validity (e.g., content and construct validity) should be examined and confirmed (Bauer *et al.*, 2012). For this reason, the psychometric properties of VR-EAL were assessed in Chapter 5. In Chapter 5, the performance on VR-EAL tasks significantly correlated with the performance on the equivalent ecologically valid tests, which also supported the construct and content validity of VR-EAL to assess these everyday cognitive functions. Also, in the same chapter, the tasks of VR-EAL were rated by participants as substantially more ecologically valid than the corresponding tasks of these tests, which may be attributed to the benefits of using immersive VR methods.

Overall, in Chapter 5, the content, construct, and ecological validity were explored and supported (Kourtesis *et al.*, 2020a). Additionally, in the same chapter, the VR-EAL showed good internal consistency (i.e., reliability; Kourtesis *et al.*, 2020a).

Furthermore, since the VR-EAL has a standardised and automated scoring method means that there are not any differences across diverse end-users (i.e., the VR-EAL has an impeccable inter-rater reliability). In addition, since the VR-EAL does not have alternate forms, the alternate-form reliability was not examined and the test-retest consistency was not examined in Chapters 4 and 5. Nonetheless, the test-retest reliability of VR-EAL should be explored in future work. However, both the validity and reliability of a test are not unitary psychometric properties, and they should be re-examined as populations and the testing context changes over time (Nunnally & Bernstein, 1994). Notably, the eventual aim of CNADs, such as VR-EAL, is their utilization for research and clinical purposes in healthy aging and clinical groups such as dementias (Anderson & Craik, 2017), attention-deficit/hyperactivity disorder and autism (Karalunas *et al.*, 2018), mild cognitive impairment (Schmitter-Edgecombe *et al.*, 2009), acquired and traumatic brain injuries (Groot *et al.*, 2002), HIV (Woods *et al.*, 2008), schizophrenia (Twamley *et al.*, 2008), and Parkinson's disease (Pirogovsky *et al.*, 2012). Thus, the administration of VR-EAL in healthy aging and clinical populations may highlight its clinical utility through an exploration of its diagnostic validity (e.g., in the detection of mild cognitive impairment) and predictive validity (e.g., predicting everyday functionality and the independence of older adults).

One limitation of my PhD work was that the VR-EAL was only administered to healthy young adults (18 – 45 years old) who are unlikely to demonstrate any cognitive impairments or disorders (Chaytor & Schmitter-Edgecombe, 2003). Hence, the validity of VR-EAL should also be studied in older adults. As an ecologically valid test, the VR-EAL may also elucidate issues associated with cognitive ageing. For example, regarding prospective memory and older adults, an age-related paradox is observed, where older adults appear impaired on laboratory-based prospective memory tasks, while they perform better than younger adults on naturalistic tasks (Schnitzspahn, Ihle, Henry, Rendell, & Kliegel, 2011). Due to their increased life-experience and crystallised intelligence, older adults appear to be more effective in using environmental cues and compensatory strategies such as having a structured plan of action (e.g., noting down the sequence of necessary tasks), setting reminders (e.g., using notes, alarm clocks, or smartphones), making stronger and more complex associations between a task and an environmental cue (e.g., seeing a building, which used to be a post-office in the past, may remind the intention to mail a postcard to a relative), and using specialised items (e.g., using a dosette box to manage medications; Chaytor & Schmitter-Edgecombe, 2003; Marsh, Hicks, & Landau, 1998; Schnitzspahn *et al.*, 2011). However, the utilisation of such techniques is not feasible in non-ecologically valid tests because of their structured procedures which only allow participants to respond or perform the task in a certain way (e.g., pressing a button on the keyboard, when seeing a specific item on the screen; Marsh *et al.*, 1998; Schnitzspahn *et al.*, 2011).

Consequently, the prospective memory age-related paradox highlights the importance of ecological validity in the assessment of everyday cognitive functioning (Chaytor & Schmitter-Edgecombe, 2003; Schnitzspahn *et al.*, 2011). However, as discussed in Chapter 1, tasks performed in the real world (e.g., Marsh *et al.*, 1998) cannot be standardized to allow their administration in other clinics or laboratories (Parsons, 2015). Also, they may not be appropriate for some individuals in challenging populations (e.g., a disabled patient using a wheelchair), they are time-consuming and expensive (e.g., they require participant transport and consent from local businesses), and they do not have experimental control over the external situation (Parsons, 2015). In contrast, immersive VR CNADs like VR-EAL enable an adequate level of experimental control, while they are more cost-effective and inclusive than real-world tasks (i.e., naturalistic tasks; Parsons, 2015).

Potentially the VR-EAL could be used in the future to investigate the age-related paradox in prospective memory functioning in older and younger adults, which may also clarify the veridicality and predictive validity of VR-EAL by examining the existence of potential relationships between the VR-EAL scores and established questionnaires assessing the ability to perform instrumental activities of daily life. Also, the inclusion of patients with mild cognitive impairment, which is a challenging population for diagnostic cognitive tests (i.e., tests frequently fail to achieve an adequately high sensitivity and specificity in differentiating individuals with mild cognitive impairment from healthy controls; Schmitter-Edgecombe *et al.*, 2009), in the same future study may inform on the predictive validity of VR-EAL by examining its sensitivity and specificity in differentiating older adults with mild cognitive impairment from healthy older adults. In summary, in line with the guidelines of AACN and NAN on providing evidence for a CNAD's utility (i.e., psychometric properties), the studies of this thesis have demonstrated the ecological validity of VR-EAL, as well as its content and construct validity in young adults. However, the experimental and clinical utility of VR-EAL should be further explored in healthy older adults and dementias (e.g., mild cognitive impairment).

7.3.4. Examinee issues (point 6)

Another important concern of AACN and NAN regarding the implementation of CNAD is that individual differences (e.g., age, culture, education, motor abilities, and computer skills) may affect the examinees' performance on CNADs (Bauer *et al.*, 2012). For these reasons, the developers of CNADs should investigate how diverse cultural, age, and educational background may affect the performance of the examinees, and then provide normative data correspondingly (Bauer *et al.*, 2012). Furthermore, cognitive, motor, or sensory disabilities might have an impact on the examinees' ability to perform the CNAD's tasks effectively; hence, the suitability of the tests for individuals with disabilities should be explored and documented (Bauer *et al.*, 2012). Finally, competency and familiarity with computers may also affect the validity of the CNAD's results (Bauer *et al.*, 2012). Indeed, there are significant individual differences pertaining to the competency and familiarity with computer use (Iverson, Brooks, Ashton, Johnson, & Gualtieri, 2009). For example, gamers have been found to have faster perceptual processing speed compared to non-gamers, regardless their performance on tasks (e.g., number of errors and correct responses; Kowal, Toth, Exton, & Campbell, 2018). Importantly, the results from computerized against paper-and-pencil tests may be substantially different in computer-familiarised against computer-naive populations (Iverson *et al.*, 2009; Feldstein *et al.*, 1999).

However, the examinee's competency in using computers mainly influences the performance on non-immersive CNADs. The user interface and procedure of non-immersive CNADs can be challenging for individuals without gaming background or familiarization with computers (Parsons *et al.*, 2018; Zaidi, Duthie, Carr, & Maksoud, 2018), especially for older adults (Werner & Korczyn, 2012; Zygouris & Tsolaki, 2015). On the other hand, immersive VR CNADs appear to rely significantly less on gaming or computing ability than non-immersive CNADs (Bohil, Alicea, & Biocca, 2011; Parsons, 2015; Teo *et al.*, 2016). The first-person perspective in conjunction with naturalistic interactions (i.e., close to real-life actions) assist non-gamers to perform comparable to gamers in immersive VR environments (Zaidi *et al.*, 2018). Indeed, the findings of the studies in this thesis indicated that the gaming ability of the examinee does not affect the utilisation of immersive VR technologies and performance on the VR-EAL. In Chapter 3, there was no significant difference between gamers and non-gamers in the duration of the VR session. Similarly, in Chapters 4 and 5, the performance on VR-EAL appeared to demonstrate no difference between gamers and non-gamers. Finally, the performance on VR-EAL was not found to be affected by the age or educational background. Therefore, the VR-EAL appears to be appropriate for the assessment of young individuals regardless of their educational background, age, or competency in using computers.

However, as discussed above, the VR-EAL should also be administered to older adults to allow an investigation of their attitudes towards VR-EAL, and whether their competency in computers affects their performance on the VR-EAL. Nevertheless, recent studies have found that older adults, after using immersive software, expressed a very positive attitude towards immersive VR technologies and rated immersive VR software as a highly pleasant experience, while they did not experience adverse VRISE (Appel *et al.*, 2020; Brown, 2019; De Vries, Van Dieën, Van Den Abeele, & Verschueren, 2018; Huygelier, Schraepen, van Ee, Abeele, & Gillebert, 2019). Also, the application of immersive VR software was feasible in older adults with lower-motor disabilities (Appel *et al.*, 2020; Brown, 2019), as well as in older adults with various levels of cognitive impairments (i.e., mild, moderate, and severe; Appel *et al.*, 2020). However, older adults were found to prefer and perform better on immersive VR software that have ergonomic and naturalistic interactions (De Vries *et al.*, 2018). Furthermore, both younger and older adults showed an increased motivation to perform cognitive tasks in immersive VR rather than traditional paper-and-pencil tests (Corriveau Lecavalier, Ouellet, Boller, & Belleville, 2020). Finally, the performance of both younger and older adults on episodic memory tasks in an immersive VR CNAD were analogous to their performance on traditional paper-and-pencil episodic memory tests, indicating that the performance of both younger and older adults was not affected by their competency in using computers (Corriveau Lecavalier *et al.*, 2020).

In this thesis, the VR-EAL, which provides ergonomic and naturalistic interactions, was rated as a highly pleasant testing experience by younger adults, whose performance on the VR-EAL was substantially correlated with their performance on equivalent paper-and-pencil tests. Therefore, based on the findings of the aforementioned studies (i.e., Appel *et al.*, 2020; Brown, 2019; Corriveau Lecavalier *et al.*, 2020; De Vries *et al.*, 2018), in conjunction with the findings of this thesis, it may be hypothesised that the future implementation of VR-EAL in older adults with diverse functionality (i.e., healthy individuals, individuals with cognitive impairments and/or lower-motor disabilities) is feasible. Furthermore, comparably to the findings of this thesis, the implementation of VR-EAL in older adults is expected to offer a pleasant testing experience without VRISE and show equivalent psychometric properties regardless their gaming/computing ability. Nevertheless, future implementations of the VR-EAL in diverse populations will allow a scrutiny of VR-EAL's psychometric properties, strengths, and limitations, which will allow the creation of more detailed documentation to assist VR-EAL's end-users with implementing VR-EAL competently.

7.4. Limitations and future directions

This thesis has also some limitations that should be considered. As the current thesis aimed to explore the appropriateness and utility of the immersive VR methods in cognitive neuroscience and neuroscience, the various types of VR software (i.e., including VR-EAL) were only administered to younger healthy adults with a relatively medium to high education. While the performance on VR-EAL was found not to be affected by age, education, or gaming ability, it would be important for these relationships to be examined in a more education- and age-diverse population including older adults. Also, as discussed above, the clinical and experimental utility of VR-EAL should be further investigated in dementia-related conditions such a mild cognitive impairment.

Furthermore, the VR-EAL as a CNAD presented some limitations in the studies of this thesis. As discussed in Chapter 4 and above, the VR-EAL induces strong placement and plausibility illusions, though, the embodiment illusion is only of moderate strength because it relies only on hands/controllers' movements. The embodiment illusion pertains to the illusion of owning a virtual body (i.e., virtual avatar; Maister *et al.*, 2015; Pan & Hamilton, 2018). The embodiment illusion is important for acquiring cognitive and behavioural data, which resemble the individual's cognition and behaviour in real life (Maister *et al.*, 2015; Pan & Hamilton, 2018). Despite this limitation, VR-EAL was rated as very similar to real life (i.e., enhanced ecological validity); however, the improvement of the embodiment illusion would probably increase the already enhanced ecological validity of VR-EAL. The most common and easy technique to create a responsive virtual body is the utilisation of software development kits (e.g., VRIK, Final IK, and IK for VR) which offer reliable and accurate inverse kinematics (i.e., animating the virtual avatar with respect to the user's movements; Lugrin *et al.*, 2018).

Nonetheless, the virtual body should be as close as possible to the actual appearance and body of the examinee. Owning a virtual body that is dissimilar to the examinee's body may affect the performance of the examinee, either positively or negatively (Maister *et al.*, 2015; Pan & Hamilton, 2018). For example, owning a virtual body which resembles that of Albert Einstein was found to significantly increase cognitive performance (Banakou, Kishore, & Slater, 2018). Regardless of whether the impact of the virtual body on cognitive performance is positive or negative, the virtual body that an immersive VR CNAD like VR-EAL offers should be as similar as possible to the examinee's body in order that the observed cognitive performance relates to the examinee's everyday cognitive ability. Hence, a future version of VR-EAL should include an application (e.g., using a photograph(s) of the examinee) which generates a virtual avatar looking similar to the participant as happens in other immersive VR software (e.g., EngageVR).

Also, as discussed in Chapter 6, the VR-EAL does not have a mechanism to measure time monitoring, although there is a digital watch which the examinee uses to monitor time. Time monitoring has been found crucial in time-based prospective memory functioning (McFarland & Glisky, 2009; Mioni & Stablum, 2014; Vanneste, Baudouin, Bouazzaoui, & Taconnat, 2016). As mentioned in Chapter 4, the VR-EAL includes a visual attention task, where the participant detects visual targets in the environment. The measurement of performance is facilitated by a gaze interaction system which uses an invisible ray emitted (i.e., ray-casting) from the forehead point between the eyes (i.e., the upper point of the nose) and straight forward to the centre of the participant's field of view. Thus, the same gaze interaction system may be implemented to accurately quantify time monitoring in terms of when and how many times the participant reads the time on the digital watch. The inclusion of a quantified time monitoring score in a future version of the VR-EAL will facilitate a more comprehensive assessment of time-based prospective memory.

Moreover, as discussed in Chapter 6, prospective memory components (e.g., retrospective), cue attributes (e.g., focality and salience of the cue) and the role of executive functions are important in prospective memory functioning, and they should be further explored in future studies. However, ecological valid CNADs like VR-EAL, which simulate everyday tasks (e.g., cooking), may be susceptible to confounding factors and fail to thoroughly examine a specific cognitive process. On the other hand, laboratory tasks are able to exclude confounding factors and permit the examination of a specific cognitive process. Nevertheless, laboratory tasks in their current form suffer from limitations such as the two-dimensional environment, the non-naturalistic and non-ergonomic responses (i.e., using a keyboard, a button box, or joystick), static stimuli, and a substantial divergence from looking realistic. The utilization of immersive VR technologies may be capable of resolving the limitations of laboratory tasks. Immersive VR laboratory experiments would be advanced further by having a

360° testing environment, which incorporates realistic and dynamic stimuli, where the participant can interact in an ergonomic and naturalistic way (i.e., using wands/controllers or her/his own hands). Therefore, immersive VR laboratory experiments would minimize the divergence from real-life conditions, while facilitating a meticulous examination of the prospective memory components and cue attributes, as well as the role of executive functions in prospective memory functioning.

As mentioned in Chapters 4 and 5, the VR-EAL can be utilized as an entire scenario for the assessment of everyday prospective memory, episodic memory, visual attention, visuospatial attention, auditory attention, and executive functions. However, the VR-EAL also offers a shorter scenario, where the aforementioned cognitive functions can be assessed except for prospective memory and auditory attention. Moreover, the tasks of VR-EAL may be administered independently (i.e., generic tutorial, the specific tutorial for this task, and the storyline task) for the assessment of a specific cognitive function (e.g., multitasking, visual attention). Hence, VR-EAL could be considered as a mini library of immersive VR assessments. In the future, the sum of VR-EAL assessments (i.e., whole scenario, short scenario, and independent tasks) in conjunction with any future immersive VR CNAD may form an open access and source library of immersive VR software for cognitive neuroscience and neuropsychology. Considering the widespread adoption of open access and source tools such as PsychoPy (Peirce, 2007; 2009), OpenSesame (Mathôt, Schreij, & Theeuwes, 2011), R software (Culpepper & Aguinis, 2011), and Psych Package (Revelle, 2011) in the last decade, the creation of an open access and source library of immersive VR software will promote the adoption of immersive VR technologies in cognitive neuroscience and neuropsychology.

Immersive VR software for implementation in cognitive neuroscience should incorporate neuroscientific methods in the future. Immersive VR technologies (i.e., HMDs) are compatible with electroencephalography (EEG; Teo *et al.*, 2016), eye-tracking (Pettersson *et al.*, 2018), and near-infrared spectroscopy (Teo *et al.*, 2016), albeit they have scarcely been implemented in combination with immersive VR (Pettersson *et al.*, 2018; Teo *et al.*, 2016). Eye-tracking may offer a detailed map with the trajectories of the examinee's eye gaze alongside with the times (i.e., the time that the examinee's gaze fell on this point) and performance on the immersive VR CNAD (Petterson *et al.*, 2018). For example, combining eye-tracking with an immersive VR CNAD may assist with clarifying whether impaired performance on a cognitive task (e.g., of abstract reasoning) is indeed due to an impaired ability on the assessed cognitive function or due to impaired attentional processes (Petterson *et al.*, 2018). Comparably to traditional approaches, combining neuroimaging techniques (e.g., EEG) with an immersive VR CNAD may inform on which brain regions are activated when a cognitive task is performed (Teo *et al.*, 2016). Also, the combined implementation of immersive VR software with neuroimaging techniques such as EEG facilitate the utilisation of a brain computer interface (BCI; i.e., a direct communication pathway between the brain and an external device), where the examinee controls her/his virtual body in the virtual environment by activating predefined brain regions (Teo *et al.*, 2016). For example, the examinee thinks the word "forward" to move her/his virtual body forward in the virtual environment. Using a BCI allows examinees with severe motor disabilities (e.g., tetraplegic) to perform the tasks in an immersive VR CNAD (Teo *et al.*, 2016). Hence, an open access and source library for immersive VR software in cognitive neuroscience should facilitate and/or incorporate some of the aforementioned neuroscientific methods (e.g., eye-tracking and EEG).

7.5. Conclusions

This thesis endeavoured to address the shortcomings pertaining to the implementation of immersive VR technologies in cognitive neuroscience and neuropsychology by providing essential technological knowledge for the selection of appropriate hardware (i.e., HMDs, external, and computer) and software, as well as guidelines for the in-house and cost-effective development of immersive VR software. In addition, an advancement of the current available immersive VR research methods was attempted by developing and validating the VRNQ and VR-EAL. The VRNQ appears to be a valid and reliable tool for the appraisal of the intensity of VRISE and the VR software features which are crucial for the alleviation or avoidance of VRISE. The VR-EAL is the first immersive VR neuropsychological battery with enhanced ecological validity for the assessment of everyday cognitive functions, which facilitates a pleasant testing experience without inducing VRISE. The VR-EAL was also found able to contribute to the theoretical framework of prospective memory, which provides further evidence for the utility of immersive VR methods in cognitive neuroscience and neuropsychology. It is hoped that the findings of these series of studies have demonstrated the utility of immersive VR methods for improving the ecological validity and realism of neuropsychological assessment.

References

Allen, C. L., & Bayraktutan, U. (2009). Oxidative stress and its role in the pathogenesis of ischaemic stroke. *International Journal of Stroke*, 4(6), 461-470.

Almeida, A., Rebelo, F., Noriega, P., & Vilar, E. (2017, July). Virtual Reality Self Induced Cybersickness: An Exploratory Study. In *International Conference on Applied Human Factors and Ergonomics* (pp. 26-33). Springer, Cham.

Anderson, N. D., & Craik, F. I. (2017). 50 years of cognitive aging theory. *The Journals of Gerontology: Series B*, 72(1), 1-6.

Anderson, F. T., McDaniel, M. A., & Einstein, G. O. (2017). Remembering to remember: An examination of the cognitive processes underlying prospective memory. *Learning and Memory: A Comprehensive Reference E*, 2, 451-463.

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., ... & Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. *Nature*, 501(7465), 97.

Anthes, C., García-Hernández, R. J., Wiedemann, M., & Kranzlmüller, D. (2016, March). State of the art of virtual reality technology. In *2016 IEEE Aerospace Conference* (pp. 1-19). IEEE.

American Academy of Clinical Neuropsychology. (2007). American Academy of Clinical Neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation. *Clinical Neuropsychology*, 21 (2), 209–231.

American Psychological Association. (2007). Record keeping guidelines. *American Psychologist*, 62, 993–1004.

APA. (2010). Ethical Principles of Psychologists and Code of Conduct (2010 Amendments). Retrieved from http://www.apa.org/ethics/code/index.aspx.

Appel, L., Appel, E., Bogler, O., Wiseman, M., Cohen, L., Ein, N., ... & Campos, J. L. (2020). Older adults with cognitive and/or physical impairments can benefit from immersive virtual reality experiences: a feasibility study. *Frontiers in Medicine*, 6, 329.

Arafat, I. M., Ferdous, S. M. S., & Quarles, J. (2018, March). Cybersickness-Provoking Virtual Reality Alters Brain Signals of Persons with Multiple Sclerosis. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 1-120). IEEE.

Arbuckle, J. L. (2014). Amos (Version 23.0) [Computer Program]. Chicago: IBM SPSS.

Arrieux, J. P., Cole, W. R., & Ahrens, A. P. (2017). A review of the validity of computerized neurocognitive assessment tools in mild traumatic brain injury assessment. *Concussion*, 2(1), CNC31.

Azzopardi, B., Auffray, C., & Kermarrec, C. (2017). Paradoxical effect of aging on laboratory and naturalistic time-based prospective memory tasks. Role of executive functions. *Canadian Journal on Aging/La Revue canadienne du vieillissement*, 36(1), 30–40.

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. *The Quarterly Journal of Experimental Psychology*, 63(4), 646-652.

Banakou, D., Kishore, S., & Slater, M. (2018). Virtually being Einstein results in an improvement in cognitive Task performance and a decrease in age bias. *Frontiers in Psychology*, 9, 917.

Banville, F., Nolin, P., Lalonde, S., Henry, M., Dery, M. P., & Villemure, R. (2010). Multitasking and prospective memory: can virtual reality be useful for diagnosis?. *Behavioural Neurology*, 23(4), 209-211.

Barberia, I., Oliva, R., Bourdin, P., & Slater, M. (2018). Virtual mortality and near-death experience after a prolonged exposure in a shared virtual reality may lead to positive life-attitude changes. *PloS One*, 13(11), e0203358.

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *Archives of Clinical Neuropsychology*, 27(3), 362-373.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., ... & Cesarini, D. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6.

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, 39(1), 15-22.

Bioulac, S., Lallemand, S., Rizzo, A., Philip, P., Fabrigoule, C., & Bouvard, M. P. (2012). Impact of time on task on ADHD patient’s performances in a virtual classroom. *European Journal of Paediatric Neurology*, 16(5), 514-521.

Bland, M. (2015). *An introduction to medical statistics*. Oxford, UK: Oxford University Press.

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

Borges, M., Symington, A., Coltin, B., Smith, T., & Ventura, R. (2018, October). HTC Vive: Analysis and Accuracy Improvement. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 2610-2615). IEEE.

Borrego, A., Latorre, J., Alcaniz, M., and Llorens, R. (2018). Comparison of oculus rift and HTC vive: feasibility for virtual reality-based exploration, navigation, exergaming, and rehabilitation. *Games Health Journal*. 7, 151–156.

Botella, C., García-Palacios, A., Villa, H., Baños, R. M., Quero, S., Alcañiz, M., & Riva, G. (2007). Virtual reality exposure in the treatment of panic disorder and agoraphobia: A controlled study. Clinical Psychology & Psychotherapy: *An International Journal of Theory & Practice*, 14(3), 164-175.

Bottari, C., Dassa, C., Rainville, C., & Dutil, É. (2010). A generalizability study of the instrumental activities of daily living profile. *Archives of Physical Medicine and Rehabilitation*, 91(5), 734-742.

Bottari, C., Shun, P. L. W., Le Dorze, G., Gosselin, N., & Dawson, D. (2014). Self-generated strategic behavior in an ecological shopping task. *American Journal of Occupational Therapy*, 68(1), 67-76.

Bouchard, S., Robillard, G., Renaud, P., & Bernier, F. (2011). Exploring new dimensions in the assessment of virtual reality induced side effects. *Journal of Computer and Information Technology*, 1(3), 20-32.

Bourdin, P., Barberia, I., Oliva, R., & Slater, M. (2017). A virtual out-of-body experience reduces fear of death. *PloS One*, 12(1), e0169343.

Bowman, M. L. (1996). Ecological validity of neuropsychological and other predictors following head injury. *The Clinical Neuropsychologist*, 10(4), 382-396.

Boyle, P. A., Cohen, R. A., Paul, R., Moser, D., & Gordon, N. (2002). Cognitive and motor impairments predict functional declines in patients with vascular dementia. *International Journal of Geriatric Psychiatry*, 17 (2), 164–169.

Bozgeyikli, E., Raij, A., Katkoori, S., & Dubey, R. (2016, October). Point & teleport locomotion technique for virtual reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (pp. 205-216). ACM.

Brade, J., Dudczig, M., & Klimant, P. (2018, September). Using Virtual Prototyping Technologies to Evaluate Human-Machine-Interaction Concepts. *In aw&I Conference* (Vol. 3).

Brandimonte, M. A., Einstein, G. O., & McDaniel, M. A. (2014). *Prospective memory: Theory and applications*. Psychology Press.

Brennesholtz, M. S. (2018, May). 3-1: Invited Paper: VR Standards and Guidelines. In *SID Symposium Digest of Technical Papers* (Vol. 49, No. 1, pp. 1-4).

Brown, J. A. (2019). An exploration of virtual reality use and application among older adult populations. *Gerontology and Geriatric Medicine*, 5.

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society*, 4(6), 547-558.

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Laure, M. C., Dawson, D. R., ... & 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.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Confidence and precision increase with high statistical power. *Nature Reviews Neuroscience*, 14(8), 585.

Cahn-Weiner, D. A., Farias, S. T., Julian, L., Harvey, D. J., Kramer, J. H., Reed, B. R., & Chui, H. (2007). Cognitive and neuroimaging predictors of instrumental activities of daily living. *Journal of the International Neuropsychological Society*, 13 (05), 747–757.

Canty, A. L., Fleming, J., Patterson, F., Green, H. J., Man, D., & Shum, D. H. (2014). Evaluation of a virtual reality prospective memory task for use with individuals with severe traumatic brain injury. *Neuropsychological Rehabilitation*, 24(2), 238-265.

Caputo, F., Greco, A., Egidio, D. A., Notaro, I., & Spada, S. (2017, July). A preventive ergonomic approach based on virtual and immersive reality. In *International Conference on Applied Human Factors and Ergonomics* (pp. 3-15). Springer, Cham.

Carlozzi, N. E., Gade, V., Rizzo, A. S., & Tulsky, D. S. (2013). Using virtual reality driving simulators in persons with spinal cord injury: three screen display versus head mounted display. Disability and Rehabilitation: *Assistive Technology*, 8(2), 176-180.

Cernich, A. N., Brennana, D. M., Barker, L. M., & Bleiberg, J. (2007). Sources of error in computerized neuropsychological assessment. *Archives of Clinical Neuropsychology*, 22S, S39–S48.

Chandra, K. (2012). A comparison of objective-C and C#: implementing Cbox class. *Journal of Computing Sciences in Colleges*, *27*(5), 178-186.

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.

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, 21(3), 217-227.

Christou, C. G., Michael-Grigoriou, D., Sokratous, D., and Tsiakoulia, M. (2018). “A virtual reality loop and wire game for stroke rehabilitation,” in *28th International Conference on Artificial Reality and 23rd Telexistence and Eurographics Symposium on Virtual Environments* (Limassol).

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology*, *55*(4), 584.

Cole, W. R., Arrieux, J. P., Ivins, B. J., Schwab, K. A., & Qashu, F. M. (2018). A comparison of four computerized neurocognitive assessment tools to a traditional neuropsychological test battery in service members with and without mild traumatic brain injury. *Archives of Clinical Neuropsychology*, 33(1), 102-119.

Collins, M. K., Ding, V. Y., Ball, R. L., Dolce, D. L., Henderson, J. M., & Halpern, C. H. (2018). Novel application of virtual reality in patient engagement for deep brain stimulation: A pilot study. Brain Stimulation: Basic, *Translational, and Clinical Research in Neuromodulation*.

Cona, G., Arcara, G., Tarantino, V., & Bisiacchi, P. S. (2012). Electrophysiological correlates of strategic monitoring in event-based and time-based prospective memory. *PLoS One*, 7(2).

Cona, G., Scarpazza, C., Sartori, G., Moscovitch, M., & Bisiacchi, P. S. (2015). Neural bases of prospective memory: a meta-analysis and the "Attention to Delayed Intention" (AtoDI) model. *Neuroscience & Biobehavioral Reviews*, 52, 21-37.

Constantinou, M., Bauer, L., Ashendorf, L., Fisher, J. M., & McCaffrey, R. J. (2005). Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tasks? *Archives of Clinical Neuropsychology*, 20, 191–198.

Conte, A. M., & McBride, D. M. (2018). Comparing time-based and event-based prospective memory over short delays. *Memory*, 26(7), 936-945.

Corriveau Lecavalier, N., Ouellet, É., Boller, B., & Belleville, S. (2020). Use of immersive virtual reality to assess episodic memory: a validation study in older adults. *Neuropsychological Rehabilitation*, 30(3), 462-480.

Costello, P. J. (1997). Health and safety issues associated with virtual reality: a review of current literature (pp. 1-23). *Advisory Group on Computer Graphics*.

Cox, D. R., & Donnelly, C. A. (2011). *Principles of applied statistics*. Cambridge University Press.

Craik, F. I., & Bialystok, E. (2006). Planning and task management in older adults: Cooking breakfast. *Memory & Cognition*, 34(6), 1236-1249.

Crovitz, H. F., & Daniel, W. F. (1984). Measurements of everyday memory: Toward the prevention of forgetting. *Bulletin of the Psychonomic Society*, 22(5), 413–414.

Culpepper, S. A., & Aguinis, H. (2011). R is for revolution: A cutting-edge, free, open source statistical package. *Organizational Research Methods*, 14(4), 735-740.

Davis, S., Nesbitt, K., & Nalivaiko, E. (2015, January). Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters. In *Proceedings of the 11th Australasian Conference on Interactive Entertainment (IE 2015)* (Vol. 27, p. 30).

de França, A. C. P., & Soares, M. M. (2017, July). Review of Virtual Reality Technology: An Ergonomic Approach and Current Challenges. In

International Conference on Applied Human Factors and Ergonomics (pp. 52-61). Springer, Cham.

D'Elia, L. F., Satz, P., Uchiyama, C. L., & White, T. (1996). *Color Trails Test: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Dennison Jr, M., D'Zmura, M., Harrison, A., Lee, M., & Raglin, A. (2019, May). Improving motion sickness severity classification through multi-modal data fusion. In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications* (Vol. 11006, p. 110060T). International Society for Optics and Photonics.

Dennison, M. S., Wisti, A. Z., & D'Zmura, M. (2016). Use of physiological signals to predict cybersickness. *Displays*, *44*, 42-52.

Detez, L., Greenwood, L. M., Segrave, R., Wilson, E., Chandler, T., Ries, T., ... & Yücel, M. (2019). A Psychophysiological and Behavioural Study of Slot Machine Near-Misses Using Immersive Virtual Reality. *Journal of Gambling Studies*, 1-16.

De Vries, A. W., Van Dieën, J. H., Van Den Abeele, V., & Verschueren, S. M. (2018). Understanding motivations and player experiences of older adults in virtual reality training. *Games for Health Journal*, 7(6), 369-376.

Dickson, P. E., Block, J. E., Echevarria, G. N., & Keenan, K. C. (2017, June). An experience-based comparison of unity and unreal for a stand-alone 3D game development course. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education* (pp. 70-75). ACM.

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, 72, 78-89.

Doherty, T. A., Barker, L. A., Denniss, R., Jalil, A., & Beer, M. D. (2015). The cooking task: making a meal of executive functions. *Frontiers in Behavioral Neuroscience*, 9, 22.

Dye, M. W., Green, C. S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science*, 18(6), 321-326.

Einstein, G. O. and McDaniel, M. A. (1996). "Retrieval processes in prospective memory: Theoretical approaches and some new empirical findings". In *Prospective memory: Theory and applications*, Edited by: Brandimonte, M., Einstein, G. and McDaniel, M. 115–141. Mahwah, NJ: Lawrence Erlbaum.

Einstein, G. O., McDaniel, M. A., Manzi, M., Cochran, B., & Baker, M. (2000). Prospective memory and aging: Forgetting intentions over short delays. *Psychology and aging*, 15(4), 671.

Einstein, G. O., Smith, R. E., McDaniel, M. A., & Shaw, P. (1997). Aging and prospective memory: The influence of increased task demands at encoding and retrieval. Psychology and aging, 12(3), 479. Farrell, S., & Lewandowsky, S.

(2010). Computational models as aids to better reasoning in psychology. *Current Directions in Psychological Science*, 19(5), 329-335.

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 & Behavior*, 4(4), 489–496.

Evans, J. J., Chua, S. E., McKenna, P. J., & Wilson, B. A. (1997). Assessment of the dysexecutive syndrome in schizophrenia. *Psychological Medicine*, 27(3), 635-646.

Factor, S. A., Podskalny, G. D., & Molho, E. S. (1995). Psychogenic movement disorders: frequency, clinical profile, and characteristics. *Journal of Neurology, Neurosurgery & Psychiatry*, 59(4), 406-412.

Farias, S. T., Park, L. Q., Harvey, D. J., Simon, C., Reed, B. R., Carmichael, O., & Mungas, D. (2013). Everyday cognition in older adults: Associations with neuropsychological performance and structural brain imaging. *Journal of the International Neuropsychological Society*, 19 (04), 430–44.

Farrell, S., & Lewandowsky, S. (2010). Computational models as aids to better reasoning in psychology. *Current Directions in Psychological Science*, *19*(5), 329-335.

Farrimond, S., Knight, R. G., & Titov, N. (2006). The effects of aging on remembering intentions: Performance on a simulated shopping task. *Applied Cognitive Psychology*, 20, 533–555.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149-1160.

Feldstein, S. N., Keller, F. R., Protman, R. E., Durham, R. L., Klebe, K. J., & Davis, H. P. (1999). A comparison of computerized and standard version of the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist*, 13, 303–313.

Ferrand, S., Alouges, F., & Aussal, M. (2017, May). Binaural Spatialization Methods for Indoor Navigation. In *Audio Engineering Society Convention* 142. Audio Engineering Society.

Figueiredo, L., Rodrigues, E., Teixeira, J., & Techrieb, V. (2018). A comparative evaluation of direct hand and wand interactions on consumer devices. *Computers & Graphics*, 77, 108-121.

Foerster, R. M., Poth, C. H., Behler, C., Botsch, M., & Schneider, W. X. (2016). Using the virtual reality device Oculus Rift for neuropsychological assessment of visual processing capabilities. *Scientific Reports*, 6, 37016.

Franzen, M. D., & Wilhelm, K. L. (1996). Conceptual foundations of ecological validity in neuropsychological assessment. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (p. 91–112). Gr Press/St Lucie Press, Inc.

Freeman, D., Bradley, J., Antley, A., Bourke, E., DeWeever, N., Evans, N., ... & Slater, M. (2016). Virtual reality in the treatment of persecutory delusions: randomised controlled experimental study testing how to reduce delusional conviction. *The British Journal of Psychiatry*, 209(1), 62-67.

Freilich, B. M., & Hyer, L. A. (2007). Relation of the repeatable batter for assessment of neuropsychological status to measures of daily functioning in dementia. *Psychological Reports*, 101 (1), 119–129.

Frommel, J., Sonntag, S., & Weber, M. (2017, August). Effects of controller-based locomotion on player experience in a virtual reality exploration game. In *Proceedings of the 12th International Conference on the Foundations of Digital Game*s (p. 30). ACM.

Gaggioli, A., Pallavicini, F., Morganti, L., Serino, S., Scaratti, C., Briguglio, M., ... & Tartarisco, G. (2014). Experiential virtual scenarios with real-time monitoring (interreality) for the management of psychological stress: a block randomized controlled trial. *Journal of Medical Internet Research*, 16(7).

Garden, S., Phillips, L. H. & MacPherson, S. E. (2001). Mid-life aging, open-ended planning, and laboratory measures of executive function. *Neuropsychology*, 15(4), 472-482.

Gavgani, A. M., Wong, R. H., Howe, P. R., Hodgson, D. M., Walker, F. R., & Nalivaiko, E. (2018). Cybersickness-related changes in brain hemodynamics: A pilot study comparing transcranial Doppler and near-infrared spectroscopy assessments during a virtual ride on a roller coaster. *Physiology & Behavior*, 191, 56-64.

Gehrke, L., Iversen, J. R., Makeig, S., & Gramann, K. (2018, September). The invisible maze task (IMT): interactive exploration of sparse virtual environments to investigate action-driven formation of spatial representations. In *German Conference on Spatial Cognition* (pp. 293-310). Springer, Cham.

George, C., Demmler, M., & Hussmann, H. (2018, April). Intelligent Interruptions for IVR: Investigating the Interplay between Presence, Workload and Attention. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (p. LBW511). ACM.

Gómez-Jordana, L. I., Stafford, J., Peper, C. L. E., & Craig, C. M. (2018). Virtual Footprints Can Improve Walking Performance in People With Parkinson's Disease. *Frontiers in neurology*, 9.

Goldstein, G. (1996). “Functional considerations in neuropsychology,” in *Ecological Validity of Neuropsychological Testing*, eds R. J. Sbordone and C. J. Long (Delray Beach, FL: GR Press/St. Lucie Press), 75–89.

Goldstein, G., McCue, M., Rogers, J., & Nussbaum, P. D. (1992). Diagnostic differences in memory test based predictions of functional capacity in the elderly. *Neuropsychological Rehabilitation*, 2(4), 307-317.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, 54, 493-503.

Gonneaud, J., Kalpouzos, G., Bon, L., Viader, F., Eustache, F., & Desgranges, B. (2011). Distinct and shared cognitive functions mediate event- and time-based prospective memory impairment in normal ageing. *Memory*, 19(4), 360–377.

Gonneaud, J., Rauchs, G., Groussard, M., Landeau, B., Mézenge, F., de La Sayette, V., ... & Desgranges, B. (2014). How do we process event-based and time-based intentions in the brain? An fMRI study of prospective memory in healthy individuals. *Human Brain Mapping*, 35(7), 3066-3082.

Gonzalez-Franco, M., & Lanier, J. (2017). Model of illusions and virtual reality. *Frontiers in Psychology*, *8*, 1125.

Goradia, I., Doshi, J., & Kurup, L. (2014). A review paper on oculus rift & project morpheus. International *Journal of Current Engineering and Technology*, 4(5), 3196-3200.

Grewe, P., Lahr, D., Kohsik, A., Dyck, E., Markowitsch, H. J., Bien, C. G., ... & Piefke, M. (2014). Real-life memory and spatial navigation in patients with focal epilepsy: ecological validity of a virtual reality supermarket task. *Epilepsy & Behavior*, 31, 57-66.

Gromala, D., Tong, X., Shaw, C., Amin, A., Ulas, S., & Ramsay, G. (2016). Mobius Floe: an Immersive Virtual Reality Game for Pain Distraction. *Electronic Imaging*, *2016*(4), 1-5.

Groot, Y. C., Wilson, B. A., Evans, J., & Watson, P. (2002). Prospective memory functioning in people with and without brain injury. *Journal of the International Neuropsychological Society*, 8(5), 645-654.

Gur, R. C., Richard, J., Hughett, P., Calkins, M. E., Macy, L., Bilker, W. B., ... & Gur, R. E. (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: standardization and initial construct validation. *Journal of Neuroscience Methods*, 187(2), 254-262.

Haines, S., Shelton, J., Henry, J., Terrett, G., Vorwerk, T., & Rendell, P. (2019, February 25). Prospective Memory and Cognitive Aging. Oxford Research Encyclopedia of Psychology. Retrieved 7 Dec. 2019, from https://oxfordre.com/psychology/view/10.1093/acrefore/9780190236557.001.0001/acrefore-9780190236557-e-381.

Hall, J. R., Vo, H. T., Johnson, L. A., Barber, R. C., & O'Bryant, S. E. (2011). The link between cognitive measures and ADLs and IADL functioning in mild Alzheimer's: What has gender got to do with it? *International Journal of Alzheimer's Disease*, 2011.

Harpe, S. E. (2015). How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*, 7(6), 836-850.

Hartanto, D., Kampmann, I. L., Morina, N., Emmelkamp, P. G., Neerincx, M. A., & Brinkman, W. P. (2014). Controlling social stress in virtual reality environments. *PloS One*, 9(3), e92804.

Hasler, B. S., Spanlang, B., & Slater, M. (2017). Virtual race transformation reverses racial in-group bias. *PloS One*, 12(4), e0174965.

Hecht, J. (2016). Optical dreams, virtual reality. *Optics and Photonics News*, 27(6), 24-31.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23 (7), 1093–1129.

Heinrichs, R. W. (1990). Current and emergent applications of neuropsychological assessment: Problems of validity and utility. *Professional Psychology: Research and Practice*, 21(3), 171.

Held, L., & Ott, M. (2018). On p-values and Bayes factors. *Annual Review of Statistics and Its Application*, 5, 393-419.

Hicks, J. L., Marsh, R. L., & Russell, E. J. (2000). The properties of retention intervals and their affect on retaining prospective memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1160–1169.

Higginson, C. I., Arnett, P. A., & Voss, W. D. (2000). The ecological validity of clinical tests of memory and attention in multiple sclerosis. *Archives of Clinical Neuropsychology*, 15(3), 185-204.

Honan, C. A., Brown, R. F., & Batchelor, J. (2015). Perceived cognitive difficulties and cognitive test performance as predictors of employment outcomes in people with multiple sclerosis. *Journal of the International Neuropsychological Society*, 21(2), 156-168.

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, 14, 332–346.

Hsieh, T. J. T., Kuo, Y. H., & Niu, C. K. (2018, July). Utilizing HMD VR to Improve the Spatial Learning and Wayfinding Effects in the Virtual Maze. *In International Conference on Human-Computer Interaction* (pp. 38-42). Springer, Cham.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.

Huygelier, H., Schraepen, B., van Ee, R., Abeele, V. V., & Gillebert, C. R. (2019). Acceptance of immersive head-mounted virtual reality in older adults. *Scientific Reports*, 9(1), 1-12.

IBM Corp. Released (2016). IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLOS Medicine*, 2(8), e124.

Ioannidis, J. P. (2014). How to Make More Published Research True. *PLOS Medicine*, 11(10), 1-6.

Ioannidis, J. P. (2015, November). Failure to replicate: sound the alarm. In *Cerebrum: The Dana forum on brain science* (Vol. 2015). Dana Foundation.

Iverson, G. L., Brooks, B. L., Ashton, V. L., Johnson, L. G., & Gualtieri, C. T. (2009). Does familiarity with computers affect computerized neuropsychological test performance? *Journal of Clinical and Experimental Neuropsychology*, 31, 594–604.

Jackson, D. L., Gillaspy Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*(1), 6.

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.

JASP Team (2017). JASP (Version 0.8.1.2) [Computer software].

Jefferson, A. L., Cahn-Weiner, D., Boyle, P., Paul, R. H., Moser, D. J., Gordon, N., & Cohen, R. A. (2006). Cognitive predictors of functional decline in vascular dementia. *International Journal of Geriatric Psychiatry*, 21 (8), 752–754.

Jenner, P. (2003). Oxidative stress in Parkinson's disease. Annals of Neurology: *Official Journal of the American Neurological Association and the Child Neurology Society*, 53(S3), S26-S38.

Jerald, J., LaViola Jr, J. J., & Marks, R. (2017, July). VR interactions. In *ACM SIGGRAPH 2017 Courses* (p. 19). ACM.

Johnson, J. K., Lui, L. Y., & Yaffe, K. (2007). Executive function, more than global cognition, predicts functional decline and mortality in elderly women. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 62 (10), 1134–1141.

Johnson, J. L. (1994). Episodic memory deficits in Alzheimer’s disease: A behaviorally anchored scale. *Archives of Clinical Neuropsychology,* 9(4), 337-346

Kane, R. L., and Parsons, T. D. (eds.). (2017). *The Role of Technology in Clinical Neuropsychology*. New York, NY: Oxford University Press.

Karalunas, S. L., Hawkey, E., Gustafsson, H., Miller, M., Langhorst, M., Cordova, M., ... & Nigg, J. T. (2018). Overlapping and distinct cognitive impairments in attention-deficit/hyperactivity and autism spectrum disorder without intellectual disability. *Journal of Abnormal Child Psychology*, 46(8), 1705-1716.

Kelly, J. W., Cherep, L. A., & Siegel, Z. D. (2017). Perceived space in the HTC vive. *ACM Transactions on Applied Perception* (TAP), 15(1), 2.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3), 203-220.

Kidder, D. P., Park, D. C., Hertzog, C., & Morrell, R. W. (1997). Prospective memory and aging: The effects of working memory and prospective memory task load. *Aging, Neuropsychology, and Cognition*, 4(2), 93-112.

Kim, A., Darakjian, N., & Finley, J. M. (2017). Walking in fully immersive virtual environments: an evaluation of potential adverse effects in older adults and individuals with Parkinson's disease. *Journal of Neuroengineering and Rehabilitation*, 14(1), 16.

Kim, H. K., Park, J., Choi, Y., & Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. Applied ergonomics, 69, 66-73.

Kim, J. W., Choe, W. J., Hwang, K. H., and Kwag, J. O. (2017). "78-2: The optimum display for virtual reality," in *SID Symposium Digest of Technical Papers*, 48, 1146–1149

Kim, K., Kim, J., Ku, J., Kim, D. Y., Chang, W. H., Shin, D. I., ... & Kim, S. I. (2004). A virtual reality assessment and training system for unilateral neglect. *Cyberpsychology & Behavior*, 7(6), 742-749.

Kim, Y. Y., Kim, H. J., Kim, E. N., Ko, H. D., & Kim, H. T. (2005). Characteristic changes in the physiological components of cybersickness. *Psychophysiology*, *42*(5), 616-625.

Kliegel, M., & Jager, T. (2006). Delayed–execute prospective memory performance: The effects of age and working memory. *Developmental neuropsychology*, 30(3), 819-843.

Kobayashi, M., Ueno, K., & Ise, S. (2015). The effects of spatialized sounds on the sense of presence in auditory virtual environments: a psychological and physiological study. *Presence: Teleoperators and Virtual Environments*, 24(2), 163-174.

Korre, D. (2019). Usability evaluation of spoken humanoid embodied conversational agents in mobile serious games. *PhD thesis*. Edinburgh, UK. University of Edinburgh.

Kortum, P., & Peres, S. C. (2014). The relationship between system effectiveness and subjective usability scores using the System Usability

Scale. *International Journal of Human-Computer Interaction*, *30*(7), 575-584.

Kourtesis, P., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2019a). Technological competence is a precondition for effective implementation of virtual reality head mounted displays in human neuroscience: a technological review and meta-analysis. *Frontiers in Human Neuroscience*, 13, 342.

Kourtesis P., Collina S., Doumas L.A.A., & MacPherson S.E. (2019b). Validation of the Virtual Reality Neuroscience Questionnaire: Maximum duration of immersive virtual reality sessions without the presence of pertinent adverse symptomatology. Frontiers in Human Neuroscience, 13, 417.

Kourtesis, P., Collina, S., Doumas, L. A., & MacPherson, S. E. (2020a). Validation of the Virtual Reality Everyday Assessment Lab (VR-EAL): An immersive virtual reality neuropsychological battery with enhanced ecological validity. *Journal of the International Neuropsychological Society*, in press.

Kourtesis, P., Korre, D., Collina, S., Doumas, L. A., & MacPherson, S. E. (2020b). Guidelines for the development of immersive virtual reality software for cognitive neuroscience and neuropsychology: the development of virtual reality everyday assessment lab (VR-EAL), a neuropsychological test battery in immersive virtual reality. *Frontiers in Computer Science*, 1, 12.

Kowal, M., Toth, A. J., Exton, C., & Campbell, M. J. (2018). Different cognitive abilities displayed by action video gamers and non-gamers. *Computers in Human Behavior*, 88, 255-262.

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, 21(9), 1148-1160.

LaViola Jr, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM Sigchi Bulletin*, *32*(1), 47-56.

LaViola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3D user interfaces: theory and practice*. Addison-Wesley Professional.

Lavoué, G., & Mantiuk, R. (2015). Quality assessment in computer graphics. In *Visual signal quality assessment* (pp. 243-286). Springer, Cham

Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103(481), 410-423.

Liu, L. L., & Park, D. C. (2004). Aging and medical adherence: The use of automatic processes to achieve effortful things. *Psychology and Aging*, 19, 318-325.

Logie, R. H., Trawley, S., & Law, A. S. (2011). Multitasking: Multiple, domain specific cognitive functions in a virtual environment. *Memory and Cognition*, 39, 1561–1574.

Lubetzky, A. V., Kary, E. E., Harel, D., Hujsak, B., & Perlin, K. (2018). Feasibility and reliability of a virtual reality oculus platform to measure sensory integration for postural control in young adults. *Physiotherapy Theory and Practice*, 1-16.

Lugrin, J. L., Ertl, M., Krop, P., Klüpfel, R., Stierstorfer, S., Weisz, B., ... & Latoschik, M. E. (2018, March). Any “body” there? avatar visibility effects in a virtual reality game. In 2*018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 17-24). IEEE.

Lum, H. C., Greatbatch, R., Waldfogle, G., & Benedict, J. (2018, September). How Immersion, Presence, Emotion, & Workload Differ in Virtual Reality and Traditional Game Mediums. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 62, No. 1, pp. 1474-1478). Sage CA: Los Angeles, CA: SAGE Publications.

Mackinlay, R. J., Kliegel, M., & Mäntylä, T. (2009). Predictors of time-based prospective memory in children. *Journal of Experimental Child Psychology*, 102(3), 251-264.

Maister, L., Slater, M., Sanchez-Vives, M. V., & Tsakiris, M. (2015). Changing bodies changes minds: owning another body affects social cognition. *Trends in Cognitive Sciences*, 19(1), 6-12.

Makatura, T. J., Lam, C. S., Leahy, B. J., Castillo, M. T., and Kalpakjian, C. Z. (1999). Standardized memory tests and the appraisal of everyday memory. *Brain Injury*. 13: 355–367.

Makeig, S., Gramann, K., Jung, T. P., Sejnowski, T. J., & Poizner, H. (2009). Linking brain, mind and behavior. *International Journal of Psychophysiology*, 73(2), 95-100.

Marsh, R. L., Hicks, J. L., & Landau, J. D. (1998). An investigation of everyday prospective memory. *Memory & Cognition*, 26(4), 633-643.Martin, B. A., Brown, N. L., & Hicks, J. L. (2011). Ongoing task delays affect prospective memory more powerfully than filler task delays. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 65(1), 48–56.

Marsh, R. L., Hicks, J. L., & Watson, V. (2002). The dynamics of intention retrieval and coordination of action in event-based prospective memory. *Journal of Experimental Psychology-Learning Memory and Cognition*, 28(4), 652–659.

Marsman, M., & Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology*, 14(5), 545-555.

Martin, B. A., Brown, N. L., & Hicks, J. L. (2011). Ongoing task delays affect prospective memory more powerfully than filler task delays. *Canadian Journal of Experimental Psychology*, 65(1), 48–56.

Martyr, A., & Clare, L. (2012). Executive function and activities of daily living in Alzheimer’s disease: A correlational meta-analysis. *Dementia and Geriatric Cognitive Disorders*, 33 (2–3), 189–203.

Matheis, R. J., Schultheis, M. T., Tiersky, L. A., DeLuca, J., Millis, S. R., & Rizzo, A. (2007). Is learning and memory different in a virtual environment?. *The Clinical Neuropsychologist*, 21(1), 146-161.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314-324.

McBride, D. M., Beckner, J. K., & Abney, D. H. (2011). Effects of delay of prospective memory cues in an ongoing task on prospective memory task performance. *Memory & Cognition*, 39(7), 1222-1231.

McBride, D. M., Coane, J. H., Drwal, J., & LaRose, S. A. M. (2013). Differential effects of delay on time-based prospective memory in younger and older adults. *Aging, Neuropsychology, and Cognition*, 20(6), 700-721

McCauley, M. E., & Sharkey, T. J. (1992). Cybersickness: Perception of self-motion in virtual environments. *Presence: Teleoperators & Virtual Environments*, 1(3), 311-318.

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 14(7), S127-S144.

McDaniel, M. A., & Einstein, G. O. (2007). *Prospective memory: An overview and synthesis of an emerging field*. Sage Publications.

McDaniel, M. A., & Scullin, M. K. (2010). Implementation intention encoding does not automatize prospective memory responding. *Memory & Cognition*, 38(2), 221-232.

McDaniel, M. A., Shelton, J. T., Breneiser, J. E., Moynan, S., & Balota, D. A. (2011). Focal and nonfocal prospective memory performance in very mild dementia: A signature decline. *Neuropsychology*, 25(3), 387.

McDaniel MA, Umanath S, Einstein GO and Waldum ER (2015) Dual pathways to prospective remembering. *Frontiers in Human Neuroscience,* 9:392.

McFarland, C. P., & Glisky, E. L. (2009). Frontal lobe involvement in a task of time-based prospective memory. *Neuropsychologia*, 47(7), 1660–1669.

McGeorge, P., Phillips, L. H., Crawford, J. R., Garden, S. E., Della Sala, S. D., Milne, A. B., . . . Callender, J. S. (2001). Using virtual environments in the assessment of executive dysfunction. *Presence: Teleoperators and Virtual Environments*, 10(4), 375–383.

Meier, B., Zimmermann, T. D., & Perrig, W. J. (2006). Retrieval experience in prospective memory: Strategic monitoring and spontaneous retrieval. *Memory*, 14, 872–889.

Meyerbroeker, K., Morina, N., Kerkhof, G. A., & Emmelkamp, P. M. G. (2013). Virtual reality exposure therapy does not provide any additional value in agoraphobic patients: a randomized controlled trial. *Psychotherapy and Psychosomatics*, 82(3), 170-176.

Milne, S., Orbell, S., & Sheeran, P. (2002). Combining motivational and volitional interventions to promote exercise participation: Protection motivation theory and implementation intentions. *British Journal of Health Psychology*, 7, 163-184.

Milleville-Pennel, I., & Charron, C. (2015). Do mental workload and presence experienced when driving a real car predispose drivers to simulator sickness? An exploratory study. *Accident Analysis & Prevention*, *74*, 192-202.

Mioni, G., & Stablum, F. (2014). Monitoring behaviour in a time-based prospective memory task: The involvement of executive functions and time perception. *Memory*, 22(5), 536–552.

Mittelstaedt, J., Wacker, J., & Stelling, D. (2018). Effects of display type and motion control on cybersickness in a virtual bike simulator. *Displays*, *51*, 43-50.

Mittelstaedt, J. M., Wacker, J., & Stelling, D. (2019). VR aftereffect and the relation of cybersickness and cognitive performance. *Virtual Reality*, 23, 143–154.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100.

Mlinac, M. E., & Feng, M. C. (2016). Assessment of activities of daily living, self-care, and independence. *Archives of Clinical Neuropsychology*, 31(6), 506-516.

Montenegro, J. M. F., & Argyriou, V. (2017). Cognitive evaluation for the diagnosis of Alzheimer's disease based on Turing Test and Virtual Environments. *Physiology & Behavior*, 173, 42-51.

Moreau, G., Guay, M. C., Achim, A., Rizzo, A., & Lageix, P. (2006). The virtual classroom: An ecological version of the continuous performance test–A pilot study. *Annual Review of CyberTherapy and Telemedicine*, 4, 59-66.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2), 98-100.

Morris, R., & Ward, G. (Eds.). (2004). *The cognitive psychology of planning*. Psychology Press.

Mottelson, A., & Hornbæk, K. (2017, November). Virtual reality studies outside the laboratory. In *Proceedings of the 23rd acm symposium on virtual reality software and technology* (p. 9). ACM.

Mullet, H. G., Scullin, M. K., Hess, T. J., Scullin, R. B., Arnold, K. M., & Einstein, G. O. (2013). Prospective memory and aging: Evidence for preserved spontaneous retrieval with exact but not related cues. *Psychology and Aging*, 28(4), 910.

Nalivaiko, E., Davis, S. L., Blackmore, K. L., Vakulin, A., & Nesbitt, K. V. (2015). Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiology & Behavior*, 151, 583-590.

Nesbitt, K., Davis, S., Blackmore, K., & Nalivaiko, E. (2017). Correlating reaction time and nausea measures with traditional measures of cybersickness. *Displays*, 48, 1-8

Newman, C. G., Bevins, A. D., Zajicek, J. P., Hodges, J. R., Vuillermoz, E., Dickenson, J. M., ... & Noad, R. F. (2018). Improving the quality of cognitive screening assessments: ACEmobile, an iPad-based version of the Addenbrooke's Cognitive Examination-III. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10(1), 182-187.

Niehorster, D. C., Li, L., & Lappe, M. (2017). The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research. *i-Perception*, 8(3), 2041669517708205.

Norris, G., & Tate, R. L. (2000). The Behavioural Assessment of the Dysexecutive Syndrome (BADS): Ecological, concurrent and construct validity. *Neuropsychological Rehabilitation*, 10(1), 33-45.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill

Oagaz, H., Schoun, B., Pooji, M., & Choi, M. H. (2018). Neurocognitive assessment in virtual reality through behavioral response analysis. *IEEE journal of Biomedical and Health Informatics*, *23*(5), 1899-1910.

Palmisano, S., Mursic, R., & Kim, J. (2017). Vection and cybersickness generated by head-and-display motion in the Oculus Rift. *Displays*, 46, 1-8.

Pan, X., & Hamilton, A. F. D. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(3), 395-417.

Paraskevaides, T., Morgan, C. J. A., Leitza, J. R., Bisby, J. A., Rendell, P. G., & Curran, H. V. (2010). Drinking and future thinking. Acute effects of alcohol on prospective memory and future simulation. *Psychopharmacology*, 208, 301–308.

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.

Parsons, T. D., & Barnett, M. (2017). Virtual Apartment-Based Stroop for assessing distractor inhibition in healthy aging. *Applied Neuropsychology: Adult*, 1-11.

Parsons, T. D., & Carlew, A. R. (2016). Bimodal virtual reality stroop for assessing distractor inhibition in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 46(4), 1255-1267.

Parsons, T. D., Courtney, C. G., & Dawson, M. E. (2013). Virtual reality Stroop task for assessment of supervisory attentional processing. *Journal of Clinical and Experimental Neuropsychology*, 35(8), 812-826.

Parsons, T. D., & McMahan, T. (2017). An initial validation of the virtual environment grocery store. *Journal of Neuroscience Methods*, 291, 13-19.

Parsons, T. D., McMahan, T., & Kane, R. (2018). Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms. *The Clinical Neuropsychologist*, 32(1), 16-41.

Peck, T. C., Seinfeld, S., Aglioti, S. M., & Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and Cognition*, 22(3), 779-787.

Peer, A., Ullich, P., & Ponto, K. (2018, March). Vive Tracking Alignment and Correction Made Easy. *In 2018 IEEE Conference on Virtual Reality and 3D User Interfaces* (VR) (pp. 653-654). IEEE.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2), 8-13.

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2, 10.

Perry, R. J., & Hodges, J. R. (2000). Relationship between functional and neuropsychological performance in early Alzheimer disease. *Alzheimer Disease & Associated Disorders*, 14 (1), 1–10.

Peirce, J., Gray, J., Halchenko, Y., Britton, D., Rokem, A., & Strangman, G. (2011). *PsychoPy–A Psychology Software in Python*.

Pettersson, J., Albo, A., Eriksson, J., Larsson, P., Falkman, K. W., & Falkman, P. (2018, June). Cognitive ability evaluation using virtual reality and eye tracking. In *2018 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)* (pp. 1-6). IEEE.

Phillips, L. H., Henry, J. D., & Martin, M. (2008). Adult aging and prospective memory: The importance of ecological validity. In M. Kliegel, M. A. McDaniel, & G. O. Einstein (Eds.), *Prospective memory: Cognitive, neuroscience, developmental, and applied perspectives* (p. 161–185). Taylor & Francis Group/Lawrence Erlbaum Associates.

Pirogovsky, E., Woods, S. P., Filoteo, J. V., & Gilbert, P. E. (2012). Prospective memory deficits are associated with poorer everyday functioning in Parkinson's disease. *Journal of the International Neuropsychological Society*, 18(6), 986-995.

Plant, R. R. (2016). A reminder on millisecond timing accuracy and potential replication failure in computer-based psychology experiments: An open letter. *Behavior Research Methods*, 48(1), 408-411.

Plant, R. R., & Turner, G. (2009). Millisecond precision psychological research in a world of commodity computers: New hardware, new problems?. *Behavior Research Methods*, 41(3), 598-614.

Plouzeau, J., Paillot, D., Chardonnet, J. R., & Merienne, F. (2015). Effect of proprioceptive vibrations on simulator sickness during navigation task in virtual environment. In *International conference on artificial reality and telexistence*, Kyoto, Japan

Poole, J. H., Ober, B. A., Shenaut, G. K., & Vinogradov, S. (1999). Independent frontal-system deficits in schizophrenia: cognitive, clinical, and adaptive implications. *Psychiatry Research*, 85(2), 161-176.

Porcino, T. M., Clua, E., Trevisan, D., Vasconcelos, C. N., & Valente, L. (2017, April). Minimizing cyber sickness in head mounted display systems: design guidelines and applications. In *Serious Games and Applications for Health (SeGAH), 2017 IEEE 5th International Conference* on (pp. 1-6). IEEE.

Quinlivan, B., Butler, J. S., Beiser, I., Williams, L., McGovern, E., O'Riordan, S., ... & Reilly, R. B. (2016). Application of virtual reality head mounted display for investigation of movement: a novel effect of orientation of attention. *Journal of Neural Engineering*, 13(5), 056006.

Rabin, L. A., Burton, L. A., & Barr, W. B. (2007). Utilization rates of ecologically oriented instruments among clinical neuropsychologists. *The Clinical Neuropsychologist*, 21(5), 727-743.

Rakkolainen, I., Turk, M., & Höllerer, T. (2016, October). A superwide-FOV optical design for head-mounted displays. In *Proceedings of the Joint 26th Int. Conf. on Artificial Reality and Telexistence & the 21th Eurographics Symp. on Virtual Environments* (ICAT-EGVE 2016) (pp. 45-48).

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.

Reger, G. M., Holloway, K. M., Candy, C., Rothbaum, B. O., Difede, J., Rizzo, A. A., & Gahm, G. A. (2011). Effectiveness of virtual reality exposure therapy for active duty soldiers in a military mental health clinic. *Journal of Traumatic Stress*, 24(1), 93-96.

Reitan, R., & Wolfson, D. (1993). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation*. Tucson, AZ: Neuropsychology Press.

Revelle, W. (2011). *An Overview of the Psych Package*. Evanston, IL: Department of Psychology, Northwestern University

Riecke, B. E., Feuereissen, D., Rieser, J. J., & McNamara, T. P. (2011, May). Spatialized sound enhances biomechanically-induced self-motion illusion (vection). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2799-2802). ACM.

Rizzo, A., Difede, J., Rothbaum, B. O., Reger, G., Spitalnick, J., Cukor, J., & Mclay, R. (2010). Development and early evaluation of the Virtual Iraq/Afghanistan exposure therapy system for combat-related PTSD. *Annals of the New York Academy of Sciences*, 1208(1), 114-125.

Rizzo, A. A., Schultheis, M., Kerns, K. A., & Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychological Rehabilitation*, 14(1-2), 207-239.

Robertson, I. H., Ward, T., Ridgeway, V., and Nimmo-Smith, I. (1994). *The Test of Everyday Attention (TEA)*. Bury St. Edmunds: Thames Valley Test Company.

Rooney, B., Burke, C., Bálint, K., O'Leary, T., Parsons, T., Lee, C. T., & Mantei, C. (2017, October). Virtual reality, presence and social cognition: The effect of eye-gaze and narrativity on character engagement. In *2017 23rd International Conference on Virtual System & Multimedia (VSMM)* (pp. 1-6). IEEE.

Rose, N. S., Luo, L., Bialystok, E., Hering, A., Lau, K., & Craik, F. I. M. (2015). Cognitive processes in the Breakfast Task: Planning and monitoring. *Canadian Journal of Experimental Psychology*, 69(3), 252-263.

Rosenberg, L. (2015). The associations between executive functions' capacities, performance process skills, and dimensions of participation in activities of daily life among children of elementary school age. *Applied Neuropsychology: Child*, 4(3), 148-156.

Rothbaum, B. O., Price, M., Jovanovic, T., Norrholm, S. D., Gerardi, M., Dunlop, B., ... & Ressler, K. J. (2014). A randomized, double-blind evaluation of D-cycloserine or alprazolam combined with virtual reality exposure therapy for posttraumatic stress disorder in Iraq and Afghanistan War veterans. *American Journal of Psychiatry*, 171(6), 640-648.

Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, 47(6), 877-903.

Ruff, R. M., Marshall, L. F., Crouch, J., Klauber, M. R., Levin, H. S., Barth, J., ... & Jane, J. A. (1993). Predictors of outcome following severe head trauma: follow-up data from the Traumatic Coma Data Bank. *Brain Injury*, 7(2), 101-111.

Ruff, R. M., Niemann, H., Allen, C. C., Farrow, C. E., & Wylie, T. (1992). The Ruff 2 and 7 selective attention test: a neuropsychological application. *Perceptual and Motor Skills*, *75*(3_suppl), 1311-1319.

Sahakian, B. J., & Owen, A. M. (1992). Computerized assessment in neuropsychiatry using CANTAB: discussion paper. *Journal of the Royal Society of Medicine*, 85(7), 399.

Schmitter-Edgecombe, M., Woo, E., & Greeley, D. R. (2009). Characterizing multiple memory deficits and their relation to everyday functioning in individuals with mild cognitive impairment. *Neuropsychology*, 23(2), 168.

Schnitzspahn, K. M., Ihle, A., Henry, J. D., Rendell, P. G., & Kliegel, M. (2011). The age-prospective memory-paradox: An exploration of possible mechanisms. *International Psychogeriatrics*, 23(4), 583-592.

Schnitzspahn, K. M., Stahl, C., Zeintl, M., Kaller, C. P., & Kliegel, M. (2013). The role of shifting, updating, and inhibition in prospective memory performance in young and older adults. *Developmental Psychology*, 49(8), 1544–1553.

Scullin, M. K., & McDaniel, M. A. (2010). Remembering to execute a goal: Sleep on it! *Psychological Science*, 21, 1028–1035.

Scullin, M. K., McDaniel, M. A., Shelton, J. T., & Lee, J. H. (2010). Focal/nonfocal cue effects in prospective memory: Monitoring difficulty or different retrieval processes?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(3), 736..

Sejnowski, T. J., Koch, C., & Churchland, P. S. (1988). Computational neuroscience. *Science*, 241(4871), 1299-1306.

Seyama, J. I., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and virtual environments*, 16(4), 337-351.

Shallice, T., & Burgess, W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, 114, 727–741.

Sharples, S., Cobb, S., Moody, A., and Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): comparison of head mounted display (HMD), desktop and projection display systems. *Displays*, 29, 58–69.

Sherman, W. R., & Craig, A. B. (2018). *Understanding virtual reality: Interface, application, and design*. San Francisco, CA: Morgan Kaufmann.

Shiban, Y., Schelhorn, I., Pauli, P., & Mühlberger, A. (2015). Effect of combined multiple contexts and multiple stimuli exposure in spider phobia: a randomized clinical trial in virtual reality. *Behaviour Research and Therapy*, 71, 45-53.

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology* (London, England: 1953), 109(3), 431.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3549-3557.

Slater, M., Spanlang, B., & Corominas, D. (2010, July). Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. In *ACM Transactions on Graphics (TOG)* (Vol. 29, No. 4, p. 92). ACM.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 6(6), 603-616.

Smith, M. A., Rottkamp, C. A., Nunomura, A., Raina, A. K., & Perry, G. (2000). Oxidative stress in Alzheimer's disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1502(1), 139-144.

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: Investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(3), 347–361.

Smith, R. E., & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), 756–777.

Smith, R. E., Hunt, R. R., McVay, J. C., & McConnell, M. D. (2007). The cost of event-based prospective memory: Salient target events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 734–746.

Soper, D.S. (2019a). A-priori Sample Size Calculator for Structural Equation Models [Software]. Available from http://www.danielsoper.com/statcalc

Soper, D.S. (2019b). Post-hoc Statistical Power Calculator for Multiple Regression [Software]. Available from http://www.danielsoper.com/statcalc

Spooner, D. M., & Pachana, N. A. (2006). Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of Clinical Neuropsychology*, 21(4), 327-337.

Sportillo, D., Paljic, A., Boukhris, M., Fuchs, P., Ojeda, L., & Roussarie, V. (2017, February). An immersive Virtual Reality system for semi-autonomous driving simulation: a comparison between realistic and 6-DoF controller-based interaction. *In Proceedings of the 9th International Conference on Computer and Automation Engineering* (pp. 6-10). ACM.

Stanney, K. M., & Hale, K. S. (2014). Handbook of virtual environments: Design, implementation, and applications. *CRC Press*.

Stanney, K. M., Kennedy, R. S., & Drexler, J. M. (1997, October). Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society annual meeting* (Vol. 41, No. 2, pp. 1138-1142). Sage CA: Los Angeles, CA: SAGE Publications.

Stevens, A., Friedel, E., Mehen, G., & Merten, T. (2008). Malingering and uncooperativeness in psychiatric and psychological assessment: Prevalence and effects in a German sample of claimants. *Psychiatric Research*, 157, 191–200.

Stevens, J. R. (2017). Replicability and reproducibility in comparative psychology. *Frontiers in Psychology*, 8, 862.

Suznjevic, D., Evans, N., Lister, R., Antley, A., Dunn, G., & Slater, M. (2014). Height, social comparison, and paranoia: an immersive virtual reality experimental study. *Psychiatry Research*, 218(3), 348-352.

Suznjevic, M., Mandurov, M., & Matijasevic, M. (2017, May). Performance and QoE assessment of HTC Vive and Oculus Rift for pick-and-place tasks in VR. In *Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference* on (pp. 1-3). IEEE.

Teo, W. P., Muthalib, M., Yamin, S., Hendy, A. M., Bramstedt, K., Kotsopoulos, E., ... & Ayaz, H. (2016). Does a combination of virtual reality, neuromodulation and neuroimaging provide a comprehensive platform for neurorehabilitation?–A narrative review of the literature. *Frontiers in Human Neuroscience*, 10, 284

Toschi, N., Kim, J., Sclocco, R., Duggento, A., Barbieri, R., Kuo, B., & Napadow, V. (2017). Motion sickness increases functional connectivity between visual motion and nausea-associated brain regions. *Autonomic Neuroscience*, 202, 108-113.

Twamley, E. W., Woods, S. P., Zurhellen, C. H., Vertinski, M., Narvaez, J. M., Mausbach, B. T., ... & Jeste, D. V. (2008). Neuropsychological substrates and everyday functioning implications of prospective memory impairment in schizophrenia. *Schizophrenia Research*, 106(1), 42-49.

Vanneste, S., Baudouin, A., Bouazzaoui, B., & Taconnat, L. (2016). Age-related differences in time-based prospective memory: The role of time estimation in the clock monitoring strategy. *Memory*, 24(6), 812-825.

Veling, W., Brinkman, W. P., Dorrestijn, E., & Van Der Gaag, M. (2014). Virtual reality experiments linking social environment and psychosis: a pilot study. *Cyberpsychology, Behavior, and Social Networking*, 17(3), 191-195.

Viirre, E., Price, B. J., & Chase, B. (2014). Direct effects of virtual environments on users. In *Handbook of Virtual Environments* (pp. 527-535). CRC Press.

Vorländer, M., and Shinn-Cunningham, B. (2014). “Virtual auditory displays,” in *Handbook of Virtual Environments*, eds K. S. Hale and K. M. Stanney (Boca Raton, FL: CRC Press), 107–134.

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... & Meerhoff, F. (2018a). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25(1), 58-76.

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... & Matzke, D. (2018b). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35-57.

Warrington, E. K., & James, M. (1991). *The visual object and space perception battery*. Bury St Edmunds: Thames Valley Test Company

Weakley, A., Weakley, A. T., & Schmitter-Edgecombe, M. (2019). Compensatory strategy use improves real-world functional performance in community dwelling older adults. *Neuropsychology*, 33(8), 1121.

Wechsler, D. (1987). *Manual for the Wechsler Memory Scale-Revised*, The Psychological Corporation, San Antonio, TX.

Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: a review. *Frontiers in psychology*, 10, 158.

Wells, J. D., Campbell, D. E., Valacich, J. S., & Featherman, M. (2010). The effect of perceived novelty on the adoption of information technology innovations: a risk/reward perspective. *Decision Sciences*, 41(4), 813-843

Werner, P., & Korczyn, A. D. (2012). Willingness to use computerized systems for the diagnosis of dementia: testing a theoretical model in an Israeli sample. *Alzheimer Disease & Associated Disorders*, 26(2), 171-178.

West, R. (2011). The temporal dynamics of prospective memory: A review of the ERP and prospective memory literature. *Neuropsychologia*, 49(8), 2233–2245

West, L. K., Curtis, K. L., Greve, K. W., & Bianchini, K. J. (2011). Memory in traumatic brain injury: The effects of injury severity and effort on the Wechsler Memory Scale-III. *Journal of Neuropsychology*, 5, 114–125.

Wetzels, R., & Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, 19(6), 1057-1064.

Wilson, B. A. (1993). Ecological validity of neuropsychological assessment: Do neuropsychological indexes predict performance in everyday activities?. *Applied and Preventive Psychology*, 2(4), 209-215.

Wilson, B. A. (2005). *The Cambridge prospective memory test: CAMPROMPT*. London: Harcourt Assessment.

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J.J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome (BADS)*. Bury St. Edmunds, UK, Thames Valley Test Company.

Wilson, B. A., Cockburn, J., & Baddeley, A. (2008). *The Rivermead behavioural memory test*. Bury St Edmunds, UK: Thames Valley Test Company.

Wilson, B. A., Emslie, H., Foley, J., Shiel, A., Watson, P., Hawkins, K., et al. (2005). *The Cambridge Prospective Memory Test: CAMPROMPT*. London: Harcourt Assessment.

Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., and Burgess, P. (1998). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychological Rehabilitation,* 8, 213–228.

Woods, S. P., Iudicello, J. E., Moran, L. M., Carey, C. L., Dawson, M. S., & Grant, I. (2008). HIV-associated prospective memory impairment increases risk of dependence in everyday functioning. *Neuropsychology*, *22*(1), 110.

Yeh, S. C., Li, Y. Y., Zhou, C., Chiu, P. H., & Chen, J. W. (2018). Effects of Virtual Reality and Augmented Reality on Induced Anxiety. IEEE *Transactions on Neural Systems and Rehabilitation Engineering*.

Zaidi, S. F. M., Duthie, C., Carr, E., & Maksoud, S. H. A. E. (2018, December). Conceptual framework for the usability evaluation of gamified virtual reality environment for non-gamers. In *Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry* (p. 13). ACM.

Zarour, M., Abran, A., Desharnais, J. M., & Alarifi, A. (2015). An investigation into the best practices for the successful design and implementation of lightweight software process assessment methods: A systematic literature review. *Journal of Systems and Software*, *101*, 180-192.

Zimmer, P., Buttlar, B., Halbeisen, G., Walther, E., & Domes, G. (2018). Virtually Stressed? A refined virtual reality adaptation of the Trier Social Stress Test (TSST) induces robust endocrine responses. *Psychoneuroendocrinology*.

Zuber, S., Kliegel, M., & Ihle, A. (2016). An individual difference perspective on focal versus nonfocal prospective memory. *Memory & Cognition*, 44(8), 1192–1203.

Zuber, S., Mahy, C. E., & Kliegel, M. (2019). How executive functions are associated with event-based and time-based prospective memory during childhood. *Cognitive Development*, 50, 66-79.

Zygouris, S., & Tsolaki, M. (2015). Computerized cognitive testing for older adults: a review. *American Journal of Alzheimer's Disease & Other Dementias®*, 30(1), 13-28.

Appendices

Appendix A: Virtual Reality Neuroscience Questionnaire (VRNQ)

VIRTUAL REALITY NEUROSCIENCE QUESTIONNAIRE

Please, from 1 to 7, **circle** the response that closely represents your opinion.

SCIENZA NUOVA
UNIVERSITA DEGLI STUDI
SUOR ORSOLA
BENINCASA

THE UNIVERSITY of EDINBURGH
School of Philosophy, Psychology
and Language Sciences

User Experience

What is the level of immersion you experienced?

1	2	3	4	5	6	7
Extremely Low	**Very Low**	**Low**	**Neutral**	**High**	**Very High**	**Extremely High**

Please write below any additional comments and/or suggestions relevant to the question above:

What was your level of enjoyment of the VR experience?

1	2	3	4	5	6	7
Extremely Low	**Very Low**	**Low**	**Neutral**	**High**	**Very High**	**Extremely High**

Please write below any additional comments and/or suggestions relevant to the question above:

How was the quality of the graphics?

1	2	3	4	5	6	7
Extremely Low	**Very Low**	**Low**	**Neutral**	**High**	**Very High**	**Extremely High**

Please write below any additional comments and/or suggestions relevant to the question above:

How was the quality of the sound?

1	2	3	4	5	6	7
Extremely Low	**Very Low**	**Low**	**Neutral**	**High**	**Very High**	**Extremely High**

Please write below any additional comments and/or suggestions relevant to the question above:

How was the quality of the VR technology overall (i.e. hardware & peripherals)?

1	2	3	4	5	6	7
Extremely Low	**Very Low**	**Low**	**Neutral**	**High**	**Very High**	**Extremely High**

Please write below any additional comments and/or suggestions relevant to the question above:

Game Mechanics

How easy was to use the navigation system (e.g. teleportation) in the virtual environment?

1	2	3	4	5	6	7
Extremely Difficult	**Very Difficult**	**Difficult**	**Neutral**	**Easy**	**Very Easy**	**Extremely Easy**

Please write below any additional comments and/or suggestions relevant to the question above:

How easy was to physically move in the virtual environment?

1	2	3	4	5	6	7
Extremely Difficult	**Very Difficult**	**Difficult**	**Neutral**	**Easy**	**Very Easy**	**Extremely Easy**

Please write below any additional comments and/or suggestions relevant to the question above:

How easy was to pick up and/or place items in the virtual environment?

1	2	3	4	5	6	7
Extremely Difficult	**Very Difficult**	**Difficult**	**Neutral**	**Easy**	**Very Easy**	**Extremely Easy**

Please write below any additional comments and/or suggestions relevant to the question above:

How easy was to use items in the virtual environment?

1	2	3	4	5	6	7
Extremely Difficult	**Very Difficult**	**Difficult**	**Neutral**	**Easy**	**Very Easy**	**Extremely Easy**

Please write below any additional comments and/or suggestions relevant to the question above:

How easy was the 2-handed interaction e.g., grab the tablet with the one hand, and push the button with the other hand?

1	2	3	4	5	6	7
Extremely Difficult	**Very Difficult**	**Difficult**	**Neutral**	**Easy**	**Very Easy**	**Extremely Easy**

Please write below any additional comments and/or suggestions relevant to the question above:

In-Game Assistance

How easy was to complete the tutorial(s)?

1	2	3	4	5	6	7
Extremely Difficult	**Very Difficult**	**Difficult**	**Neutral**	**Easy**	**Very Easy**	**Extremely Easy**

Please write below any additional comments and/or suggestions relevant to the question above:

How helpful was/were the tutorial(s)?

1	2	3	4	5	6	7
Extremely Unhelpful	**Very Unhelpful**	**Unhelpful**	**Neutral**	**Helpful**	**Very Helpful**	**Extremely Helpful**

Please write below any additional comments and/or suggestions relevant to the question above:

How did you feel about the duration of the tutorial(s)?

1	2	3	4	5	6	7
Extremely More Time Needed	**Much More Time Needed**	**More Time Needed**	**Neutral**	**Enough Time Available**	**Much Time Available**	**Plenty of Time Available**

Please write below any additional comments and/or suggestions relevant to the question above:

How helpful were the in-game instructions for the task you needed to perform?

1	2	3	4	5	6	7
Extremely Unhelpful	**Very Unhelpful**	**Unhelpful**	**Neutral**	**Helpful**	**Very Helpful**	**Extremely Helpful**

Please write below any additional comments and/or suggestions relevant to the question above:

How helpful were the in-game prompts e.g. arrows showing the direction, or labels?

1	2	3	4	5	6	7
Extremely Unhelpful	**Very Unhelpful**	**Unhelpful**	**Neutral**	**Helpful**	**Very Helpful**	**Extremely Helpful**

Please write below any additional comments and/or suggestions relevant to the question above:

VR Induced Symptoms and Effects (VRISE)

Did you experience nausea?

1	2	3	4	5	6	7
Extremely Intense Feeling	**Very Intense Feeling**	**Intense Feeling**	**Moderate Feeling**	**Mild Feeling**	**Very Mild Feeling**	**Absent**

Please write below any additional comments and/or suggestions relevant to the question above:

Did you experience disorientation?

1	2	3	4	5	6	7
Extremely Intense Feeling	**Very Intense Feeling**	**Intense Feeling**	**Moderate Feeling**	**Mild Feeling**	**Very Mild Feeling**	**Absent**

Please write below any additional comments and/or suggestions relevant to the question above:

Did you experience dizziness?

1	2	3	4	5	6	7
Extremely Intense Feeling	**Very Intense Feeling**	**Intense Feeling**	**Moderate Feeling**	**Mild Feeling**	**Very Mild Feeling**	**Absent**

Please write below any additional comments and/or suggestions relevant to the question above:

Did you experience fatigue?

1	2	3	4	5	6	7
Extremely Intense Feeling	**Very Intense Feeling**	**Intense Feeling**	**Moderate Feeling**	**Mild Feeling**	**Very Mild Feeling**	**Absent**

Please write below any additional comments and/or suggestions relevant to the question above:

Did you experience instability?

1	2	3	4	5	6	7
Extremely Intense Feeling	**Very Intense Feeling**	**Intense Feeling**	**Moderate Feeling**	**Mild Feeling**	**Very Mild Feeling**	**Absent**

Please write below any additional comments and/or suggestions relevant to the question above:

Virtual Reality Neuroscience Questionnaire (VRNQ) – Scores

Section	Score	Minimum Cut-offs	Parsimonious Cut-offs
User Experience		≥ 25	≥ 30
Game Mechanics		≥ 25	≥ 30
In-Game Assistance		≥ 25	≥ 30
VRISE		≥ 25	≥ 30
Total VRNQ		≥ 100	≥ 120

The median of each sub-score and totals score should meet the suggested cut-offs to support that the evaluated VR software has an adequate quality without any significant VRISE.
The utilisation of the parsimonious cut-offs more robustly supports the suitability of the VR software.

The VR Neuroscience Questionnaire (VRNQ) was developed by Panagiotis Kourtesis
in affiliation with University of Edinburgh & University Suor Orsola Benincasa of Naples.

SCIENZA NUOVA
UNIVERSITA DEGLI STUDI SUOR ORSOLA BENINCASA

THE UNIVERSITY of EDINBURGH
School of Philosophy, Psychology and Language Sciences

Appendix B: VR-EAL Collected Data Sample

Scene - 1 - Tutorial Basic Interactions	
Tutorial Time = 125.9119	
Scene - 2 - Tutorial - Interactive Boards	
Interactive Boards Tutorial Time = 262.7656	
Scene - 3 Bedroom - Immediate Recognition & Planning	
PM Intro Notes Score = 6	Immediate Specific Memory Score = 18
Planning/Thinking Time = 6.042221	Immediate Qualitative Memory Score = 0
Immediate Recognition Time = 3.059387	Imm. Quantitative Memory Score = 0
Map Planning Time = 32.21278	Immediate Proxy Memory Score = 0
Scene Time Bedroom = 273.2998	Immediate Memory Total Score = 18
	Planning Map Score = 15
Scene - 4-Tutorial Prospective Memory	
PM Tutorial Time = 115.6898	
Scene - 5 -Tutorial Cooking	
Tutorial Cooking Time = 223.2638	
Scene - 6-Kitchen Cooking Task (Multitasking)	
Cooking Score = 11	PM Meds Score = 6
Task Time = 53.27881	Times Notes Used = 2
Scene Time = 97.93003	Time Spent Reading the Notes = 7.584717
Scene - 7-Tutorial Collect Items	
Collect Items Tutorial Time = 214.8167	
Scene - 8-LivingRoom Collect Items (Selective Visuospatial Attention)	
Visuospatial Attention Score = 6	Times Notes Used = 2
Task Living-Room Time = 39.93291	Time Spent Reading the Notes = 4.232811
Selective Attention Mistakes = 3	Scene Living-Room Time = 90.64527
PM_PIE Score = 6	
Scene - 9-Tutorial Interact With 3D Characters	
Tutorial Time = 148.2682	
Scene - 10 - PM Task (Meet Alex)	
PM Score = 4	Time Spent Reading the Notes = 18.1607
Times Notes Used = 6	Scene Time = 70.98004
Scene - 11-Tutorial Gaze Interaction	
Gaze Tutorial Total Time = 185.7015	Times of Practicing Gaze Interaction = 1
Scene - 12- On the Road (Selective Visual Attention i.e., SVA; Spot the Targets)	
SVA Score = 13	SVA Distractor (Same Shape) = 0
SVA Correct on Right = 6	SVA Distractor (Same Colour) Right = 0
SVA Correct on Left = 7	SVA Distractor (Same Colour) Left = 0
SVA Correct = 13	SVA Distractor (Same Colour) = 0
SVA Distractor (Same Shape) Right = 0	Total Time of the scene = 192.9267
SVA Distractor (Same Shape) Left= 0	
Scene - 13 -Tutorial Shopping	
Collect Items Tutorial Time = 87.6929	

Scene - 14 - Supermarket: Delayed Recognition Task	
Delayed Recognition Score = 20	Task Time = 142.6777
Specific Memory = 10	Times Notes Used = 2
Proxy Memory = 0	Time Spent Reading the Notes = 8.08563
Qualitative Distraction = 0	Scene Time = 189.6496
Quantitative Distraction = 0	
Scene - 15 - Supermarket (Outside): Prospective Memory Task	
PM Score = 6	Time Spent Reading the Notes = 13.48413
Times Notes Used = 2	Scene Time = 33.99336
Scene - 16 - Bakery: Prospective Memory Task	
PM Score = 0	Time Spent Reading the Notes = 25.3338
Times Notes Used = 3	Scene Time = 67.80582
Scene - 17 - Library: Prospective Memory Task	
PM Score = 6	Time Spent Reading the Notes = 0
Times Notes Used = 0	Scene Time = 26.94465
Scene - 18 - Auditory Attention Tutorial	
Times to Practice = 1	Tutorial Time = 181.7337
Scene - 19 – On the Road (Selective Auditory Attention)	
Scene Time = 217.0072	False Sounds = 0
Auditory Attention Scene Time = 224.241	False Sounds on The Right = 0
Auditory Attention Score = 28	False Sounds on The Left = 0
Sum of Sounds Detected = 14	Distractor - High Pitch Sounds = 0
Sounds Detected by False Controller = 0	Distractor - High Pitch Right = 0
Sounds on The Right Side, Left Cont. = 0	Distractor - High Pitch Left = 0
Sounds on The Left Side Right Cont. = 0	Distractor - Low Pitch Sounds = 0
Correct Sounds = 14	Distractor - Low Pitch Right = 0
Correct Sounds on The Right = 7	Distractor - Low Pitch Left = 0
Correct Sounds on The Left = 7	
Scene - 21 - Back Home - Prospective Memory Task	
PM Score = 3 Semantic Mistake(!) = 1	Times Notes Used = 2
Scene Time = 29.7544	Time Spent Reading the Notes = 4.968964
Scene - 22 - Final Scene - Back Home - Prospective Memory Task	
Total Time - Sorting Task = 67.71269	Times Notes Used = 0
Meds (Time) Final Scene - PM Score = 6	Time Spent Reading the Notes = 0
	Scene Time = 90.33228