Kent Academic Repository
Full text document (pdf)

Citation for published version

DOI

Link to record in KAR

Document Version

VStore: Feasibility and acceptability of a novel virtual reality functional cognition task

Lilla A. Porffy[1]*, Mitul A. Mehta[1], Elias Mouchlianitis[1,2] and Sukhi S. Shergill[1,3,4]

[1]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom, [2]School of Psychology, University of East London, London, United Kingdom, [3]Kent and Medway Medical School, Canterbury, United Kingdom, [4]Kent and Medway National Health Service and Social Care Partnership Trust, Gillingham, United Kingtom

Virtual reality (VR) is becoming an increasingly popular tool in neuroscience and mental health research. In recent years, efforts have been made to virtualise neuropsychological testing with the intent to increase the ecological validity of cognitive assessments. However, there are some limitations in the current literature—feasibility and acceptability data are often not reported or available and sample sizes have generally been small. In this study, we describe the development and establish the feasibility and acceptability of use of a novel functional cognition VR shopping task, VStore, in three separate samples with data from a total of 210 participants. Two samples include healthy volunteers between the ages of 20 and 79 and there is one clinical cohort of patients with psychosis. Main VStore outcomes were: 1) verbal recall of 12 grocery items, 2) time to collect items, 3) time to select items on a self-checkout machine, 4) time to make the payment, 5) time to order hot drink, and 6) total time. Feasibility and acceptability were assessed by the completion rate across the three studies. VR induced adverse effects were assessed pre- and post-VStore administration to establish tolerability. Finally, as an exploratory objective, VStore's ability to differentiate between younger and older age groups, and between patients and matched healthy controls was examined as preliminary indication of its potential utility. The overall completion rate across the studies was exceptionally high (99.95%), and VStore did not induce any adverse effects. Additionally, there was a clear difference in VStore performance metrics between both the patients and controls and between younger and older age groups, suggesting potential clinical utility of this VR assessment. These findings demonstrate that VStore is a promising neuropsychological tool that is well-tolerated and feasible to administer to both healthy and clinical populations. We discuss the implications for future research involving neuropsychological testing based on our experience and the contemporary literature.

Introduction

Virtual reality (VR) is becoming an increasingly popular tool in neuroscience and mental health research. A review published in 2017 identified 285 VR studies concerned with the assessment, understanding and treatment of mental health disorders (Freeman et al., 2017). In this context, VR is defined as a computerised three-dimensional (3D) representation of the world, that allows for the evaluation of real-time cognitive, emotional, behavioural, and physiological responses to an environment (Eichenberg and Wolters, 2012). This is due to its immersive and interactive properties that enables the user to manipulate dynamic perceptual stimuli, generating a psychological sensation of "being there" (i.e., a sense of presence) (Slater and Bouchard, 2004). Hence, this versatile technology can replicate the challenges and complexities of everyday life, while maintaining complete experimental control (Parsons, 2015).

Studies examining the feasibility and acceptability of VR interventions have been generally favourable. The use of VR has been most extensively studied in exposure therapy for anxiety disorders. Virtual reality exposure therapy (VRET) is efficacious compared to being on the waiting list for treatment (Cohen's d = 1.12, 95%CI = 0.71–1.52), and has similar acceptance rates to standard interventions such as cognitive behavioural and *in vivo* therapies but is not more efficacious (Cohen's d = 0.16, 95%CI = −0.03–0.36) (Opriş et al., 2012). Similar findings have been reported in psychosis. One of the earliest examples is the virtual London Underground train-carriage environment populated with passengers, which was found to be safe and well-tolerated in people with psychosis (Valmaggia et al., 2007; Fornells-Ambrojo et al., 2008). Various VR environments, designed for therapeutic purposes, have been established as feasible and acceptable in disorders including depression (Migoya-Borja et al., 2020), eating disorders (Clus et al., 2018), and substance use (Segawa et al., 2020).

Another rapidly developing frontier in VR research is neuropsychological testing. In the past decade, efforts have been made to adapt classic tasks such as the Stroop and Paced Auditory Serial Addition tests (Parsons et al., 2013; Parsons and Courtney, 2014). More recently however, the focus has shifted towards more complex, "real-world" scenarios that engage multiple cognitive functions. One specific example is the Virtual Reality Functional Capacity Assessment (VRFCAT), which has been approved for the Food and Drug Administration's (FDA) Qualification Program to assess functional capacity in schizophrenia (US Food and Drug Administration, 2020). Studies have found that the VRFCAT is a feasible assessment tool in both people with psychosis and subjective cognitive impairment, there is a significant difference in performance between patients and controls, and it is convergent with existing gold-standard measures of cognition and functional capacity (Ruse et al., 2014; Atkins et al., 2018).

The drive to virtualise neuropsychological testing may stem from the inherent limitations of standard measures. Most have low ecological validity—the degree to which performance on a test corresponds to real-world performance (Chaytor and Schmitter-Edgecombe, 2003); poor engagement, stemming from disinterest during testing, may compromise performance and thus test sensitivity (Millan et al., 2012; Fervaha et al., 2014); and some take a long time to complete and require a trained administrator. While VR assessments may remedy some of these issues, they come with their own challenges. Being fully immersed in a virtual environment through a head-mounted display (HMD) may induce cybersickness (i.e., nausea, instability, disorientation, headaches, eyestrain, and tiredness during or after immersion) (Guna et al., 2020), this in turn may reduce the sense of presence (and ecological validity) (Weech et al., 2019). However, sense of presence (and cognitive load) is the highest with HMD compared to two-dimensional (2D) and stereoscopic 3D displays when playing video games (Roettl and Terlutter, 2018).

VR assessment tools may have increased sensitivity in older adults. A review comparing clinical and healthy populations on various cognitive tasks (e.g., executive function and memory) embedded in VR found a large effect size (Hedges' g = .95) favouring healthy controls, which increased further as the participants' age increased (Negut et al., 2016). Similar findings have been reported in recent studies, showing that age was the strongest predictor of performance on the ECO-VR assessment composed of tasks related to activities of daily living (Oliveira et al., 2018); and vice versa, performance on two immersive VR assessments involving a parking simulator and chemistry laboratory were found to be better predictors of chronological age than traditional cognitive tasks (Davison et al., 2018). This indicates that these type of VR environments may be particularly suitable for assessing age-related cognitive decline, and perhaps neurodegenerative disorders. Nonetheless, other factors may contribute to these findings, such as technological proficiency, as older individuals are less likely to engage with a wide variety of technologies compared to younger adults (Olson et al., 2011).

While research into the applications of VR in the treatment and assessment of mental illness is an exciting novel field with enormous potential, there are some major gaps in the literature. Feasibility data are often not reported or available, particularly for VR environments developed for assessment (e.g., Davison et al., 2018). In addition, many VR studies have a small sample size that is not representative of the end-user (Rus-Calafell et al., 2017; Birckhead et al., 2019). The Virtual Reality Clinical Outcomes Research Experts (VR-CORE) committee have recently described the present state of therapeutic VR research as "the Wild West with a lack of clear guidelines and standards" (Birckhead et al., 2019). The expert group recommended that

early testing of VR treatments focuses on feasibility, acceptability, tolerability, and initial clinical efficacy, which could also be applied to the development of VR assessments.

In this study, we aim to describe the development and establish the feasibility and acceptability of administering a novel functional cognition VR task, VStore. Unlike other similar tools, such as the VRFCAT, VStore is a fully immersive VR assessment. Feasibility and acceptability were measured by completion rate across three separate samples—two including healthy adults, and one including patients with psychosis. In studies 2 and 3, we also probe tolerability by testing whether VStore induces any common adverse effects associated with HMD. In addition, we present descriptive statistics for main VStore outcome variables, and evaluate their distribution to summarise main features and ascertain the best methods to data transformation and analysis. We also evaluate the representativeness of the samples and comment on its implications for inference. As an additional exploratory objective, we examine VStore's ability to differentiate between age groups in studies 1 and 2, and patients with psychosis from Study 3 and gender and age matched healthy adults from Study 1 to provide a preliminary indication of its potential utilities. We present effect sizes for these analyses to enable power calculations in future research. Finally, based on these findings and the available literature, we provide some suggestions on how to enhance the quality of VR assessment research.

VStore development and specifications

Background

We developed VStore with the intention to measure multiple cognitive functions and functional capacity simultaneously during an ecologically valid shopping task. VStore has the potential to be employed for both clinical and research purposes. Therapeutic trials of cognitive-enhancing compounds have been consistently disappointing across a range of neuropsychiatric disorders including Alzheimer's disease and schizophrenia (Keefe et al., 2013; Plascencia-Villa and Perry, 2020). Most pharmacological agents have failed in phase III trials (Knopman et al., 2021), despite demonstrating positive effects in earlier phases (Sevigny et al., 2016). This raises several important questions about theoretical and methodological considerations around cognitive testing.

From a theoretical perspective, more emphasis should be placed on functional cognition—the ability to carry out everyday activities (Wolf et al., 2019)—rather than measuring changes in performance on a construct-driven cognitive battery, which often does not demonstrate a robust relationship to day-to-day functioning (Harvey et al., 2017). Indeed, regulatory authorities, such as the

FDA, have mandated the need for establishing functional improvements, alongside change in cognitive performance, as a condition for drug approval both in Alzheimer's disease and schizophrenia (Buchanan et al., 2005; Sabbagh et al., 2019). From a methodical perspective, motivated engagement is essential to complete a cognitive evaluation as performance and thus test sensitivity may be hindered by poor attention (Millan et al., 2012). Current testing procedures are often demanding, and while attention and motivation both fall off with time in otherwise healthy individuals, this is compounded in patients with cognitive impairment (Fervaha et al., 2014).

Finally, we considered the neuroanatomy of spatial navigation, and aimed to create a task that requires complex navigational strategies. Spatial navigation is inherent to virtual environments (Slater and Sanchez-Vives, 2016); therefore, the generation and storage of internal maps must be utilised to plan movement. Internal maps are generated by place and grid cells in the hippocampus and entorhinal cortex, respectively (O'Keefe et al., 1971; Hafting et al., 2005); and modulated by landmarks and environmental cues (Bermudez-Contreras et al., 2020), encoding an individual's position in space through allocentric processing (Herweg and Kahana, 2018). The basal ganglia-cortical circuit are believed to contribute to stimulus-response associations and procedural memory and may be related to route learning *via* egocentric navigation (Chersi and Burgess, 2015). VStore requires both allocentric and egocentric navigational strategies and thus likely to engage the relevant brain regions and circuits; therefore, it may potentially be suitable for testing disorders in which these functions are impaired. For example, decline in spatial navigation ability is a marker for pre-clinical Alzheimer's disease (Coughlan et al., 2018), and hippocampal degeneration is a feature of other neuropsychiatric conditions such as psychosis (Heckers, 2001).

VStore task

VStore takes approximately 30–40 min to complete; including orientation, instructions, practice, and testing. First, participants are shown a 5-minute-long instruction video explaining how to use the VR equipment and what to expect after entering VStore. The video is followed by additional verbal instructions explaining the steps required to complete the assessment (Supplementary Material S1).

Orientation and practice are set in a courtyard environment designed for VR acclimatisation. Here, participants learn how to move around through instant teleportation (see *Movement Parameters Section*) and manipulate objects using a wand controller. Instructions regarding movement is provided by the avatar standing in the courtyard (Figure 1). Object manipulation is first explained in the instruction video, then practice is guided by the researcher: "Now you have learned how to move around and teleport, can you see the Tropicana Orange Juice by the tree? Can you teleport there and practise picking it up

FIGURE 1
VStore Acclimatisation Courtyard. © Porffy et al. (2022). Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.01.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0). No changes were made in the reproduction of this article.

FIGURE 2
VStore fruit and vegetable section and the virtual avatar by the entrance. © Porffy et al. (2022). Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.01.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0). No changes were made in the reproduction of this article.

and bagging it, please?". There is no time limit for the practice phase—once participants picked up all six objects placed in the courtyard and report feeling confident with regards to movement and object manipulation, VStore is initiated. Orientation and practice normally take 5–10 min.

Testing is set in a minimarket environment depicting a fruit and vegetable section (Figure 2); six aisles of foodstuff, snacks, drinks, and toiletries; fridges with chilled drinks, sandwiches, and freezers with frozen meals. In addition, there are checkout and self-checkout counters (Figure 3), and a coffee shop is situated at the back of the shop. Sixty-five items—organised into nine categories—were created to fill the minimarket (Supplementary Material S2).

At the start, 12 items are read out from a shopping list by the avatar standing near the entrance (Figure 2). The first task of the

FIGURE 3
VStore self-checkout machine. © Porffy et al. (2022). Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.01.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0). No changes were made in the reproduction of this article.

FIGURE 4
Flowchart depicting the steps required to complete VStore, its corresponding cognitive domains, and outcome variables. © Porffy et al. (2022). Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.01.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0). No changes were made in the reproduction of this article.
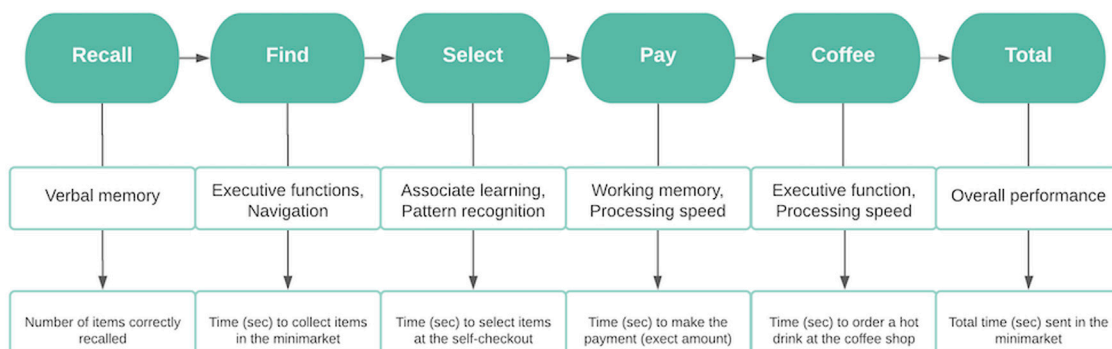
participant is to memorise and recall as many items from this list as possible. Following recall, the participant is presented with the shopping list and asked to move around the environment and collect the specified items as fast and accurate as possible, without making any mistakes such as bagging an item that is not listed. Once all the items are collected, they are required to select and pay for them at a self-checkout machine, providing the exact cost (Figure 3). The task concludes by the participant having to order a hot beverage from the coffee shop. They are unable to progress to the next phase of the assessment, without completing the previous step. If the participant gets stuck, they can approach the avatar by the entrance for instructions on the next step or ask the researcher for guidance.

Each action within the assessment maps an embedded cognitive task (e.g., recall of shopping list items measures verbal memory); while each task is assessed by performing actions that require almost identical procedures as when shopping in real life, offering a concurrent measure of functional ability. Cognitive domains engaged during the different aspects that are assessed by VStore include those commonly impaired in neuropsychiatric disorders: processing speed, verbal memory, executive function, navigation, paired-associate learning, pattern recognition, and working memory. Total time to complete VStore is treated as an overall composite score. The steps required to complete the assessment and its corresponding outcomes are summarised in Figure 4.

Software

VStore was developed using Unreal Engine 4 (UE4) written in C++ programming language. It was originally developed for Microsoft Windows operating system; but since has been adapted to PlayStation 4 and could also be adapted to other platforms. VStore is a propriety software developed in collaboration by Vitae VR Ltd. and King's College London.

Equipment specification

The VStore software was run in SteamVR on an Alienware Area-51 R² personal computer or PC (Processor: Intel® Core™ i7-6800K CPU 3.40 GHz, RAM: 32GB, Video Card: Nvidia TITAN X Pascal, OS: Microsoft Windows 10 Pro). The VR environment was displayed using HTC VIVE™ Virtual Reality System, including an HMD, two wand controllers, two tracking sensors, and a link box to connect the headset to the computer *via* HDMI and USB 2.0 cables. The HMD has a Dual AMOLED 3.6″ diagonal screen with resolution of 1080×1200 pixels per eye (2160×1200 pixels combined), refresh rate of 90 Hz, and field-of-view of 110°. This set-up allows for full immersion, providing a 360° first-person view of the virtual environment.

Movement parameters

VStore with HTC Vive set-up allows for six degrees of freedom motion tracking including rotational movement around the x, y, and *z* axes; and translational movement along those axes moving forward or backward, left or right, and up or down. Participants are free to walk around in the environment scaled to the testing room at 3.6×3.1 m. However, the testing space is smaller than the virtual minimarket environment; therefore, participants are also required to use instant teleportation. This involves pressing the trackpad on the wand controller and aiming at the desired destination, and upon release, participants teleport to that location. This method is instantaneous and does not include animation that explains the movement, which may reduce ecological validity, but it also reduces the likelihood of evoking adverse effects (Boletsis and Cedergren, 2019).

Pilot study

A pilot study was conducted among male ($n = 6$, 31.58%) and female ($n = 13$, 68.42%) college students and staff to test the usability of VStore. All participants successfully completed the task with a total average time of 560.88 s (standard deviation = 167.60, range = 278.55–963.04). However, some issues were noted during self-checkout. When participants scrolled up and down on the right-hand side of the screen to check whether they have selected all items from the shopping list, they often accidently deselected some. The screen was too small, and the scrolling buttons were situated too close to the list of items. This bug substantially increased the time spent on the VStore Select task. Therefore, the self-checkout machine was redesigned to make the screen larger, so all items were visible at once. In addition, the option to deselect individual items was removed; instead, if participants made the mistake of adding an item that was not on the shopping list, they had to clear the complete selection list and restart the process.

Study 1: Virtual reality cognitive assessment in healthy volunteers

Materials and methods

Participants

One hundred and forty-two healthy male and female participants between 20 and 79 years of age were recruited through advertisement placed in college circulars, charity newsletters, local businesses, and on social media. Participants were required to have a good standard of spoken and written English and be able to provide informed consent. Participants were excluded if they had a diagnosis of an Axis I disorder according to the Diagnostic and Statistical Manual of Mental Disorders—5th edition (DSM-5) (American Psychiatric Publishing, 2013), alcohol and/or substance use disorder, a history of severe motion sickness, a neurological illness, mobility issues, or were pregnant.

Measures
VStore

Following acclimatisation, VStore was completed as described above. The shopping list included: 1) Cornflakes, 2) Tropicana Orange Juice, 3) Coca Cola, 4) Full Fat Milk, 5) Tuna Sandwich, 6) Head and Shoulders, 7) Colgate Toothpaste, 8) Red Apple, 9) Lemon Curd, 10) Potato, 11) Orange, and 12) Brown Bread. The items were chosen to align with the list used in the pilot study.

Wechsler abbreviated scale of intelligence—1st edition

The abbreviated version of the Wechsler Adult Intelligence Scale was completed to derive an age relative intelligence quotient (IQ) (Wechsler, 1999). Specifically, the two-item scale, including matrix and vocabulary tests was used, measuring fluid and crystallised intelligence, respectively.

Technological familiarity questionnaire

We developed a short self-report questionnaire to assess technological familiarity in the sample—Supplementary Material S3 (Porffy et al., 2022). Participants were asked 13 questions to ascertain frequency, comfort, and ability in their technology use. Higher scores indicate better familiarity

with technology. The internal consistency of the questionnaire was high (Cronbach's α = 0.89, 95%CI = 0.87–0.91).

Procedures

Potential participants were pre-screened over the phone. Those eligible were invited for a single study visit lasting up to two and half hours. First, informed consent was obtained, followed by demographics, and brief mental and physical health history to confirm eligibility. Participants were asked to complete the TFQ, before progressing onto the cognitive assessments including VStore and the Cogstate computerised battery (Porffy et al., 2022). Finally, the WASI-I was completed. Participants were compensated for their time and reimbursed for travel. Ethical approval was granted by the Psychiatry, Nursing and Midwifery Research Ethics Committee, King's College London (LRS-16/17-4540).

Analysis

Feasibility and acceptability of VStore was assessed by calculating the overall completion rate—the number of participants successfully completing VStore divided by the total sample size, expressed as a percentage. Demographic characteristics of the sample were examined using means, standard deviations (SDs), and percentages, stratified by age group. Group differences in gender were tested using Chi-square test. Group differences in IQ, years spent in education (including primary, secondary, and higher) and TFQ total score were probed using one-way ANOVA and the Tukey post-hoc test where appropriate. Descriptive statistics for main VStore outcomes (see Figure 4) were calculated, and sample distributions were visually evaluated using histograms and density plots. Prior to data analysis, extreme outlier values (x −/+ 2.5 SDs) were removed. Three participants with outliers on more than one VStore outcome were eliminated from data analysis. Further nine single outlier values were removed—one from VStore Find, one from VStore Select, three from VStore Pay, and four from VStore Coffee. Deviations from the normal distribution were evaluated statistically using the Shapiro-Wilk test. Finally, as an additional exploratory objective, the presence of group differences in VStore Total Time was established using the Kruskal-Wallis test and effect size (r), which was deemed suitable based on the results of the Shapiro-Wilk test. Benjamini-Hochberg corrected Dunn test was used to explore where group differences lay. The potential confounding effects of IQ and technological familiarity were assessed using ANCOVA.

Results

VStore completion rate

Out of the 142 participants, three had missing VStore data. One participant from the 20–29 age group withdrew consent, one participant from the 50–59 age group could not complete VStore due to technical complications, and one participant from the 70–79 age group could not complete VStore due to fatigue. Counting all three exclusions, the completion rate was 97.88%.

Taking only the participant in the 70–79 age group into account, who completed VStore Find and Select (taking 37 min) before discontinuing due to exhaustion, the completion rate was 99.29%.

Sample characteristics

Demographic information—including age, gender, IQ, years spent in education, and total TFQ score—is presented in Table 1.

There were no differences between the groups in gender ratio. However, there were significant group differences in IQ score. The 70–79 age group had a higher IQ compared to both the 20–29 (mean difference [MD] = 9.07, t = 3.33, p = .016) and 40–49 (MD = 10.44, t = 3.64, p = .005) age groups. The 60–69 group also had a higher IQ on average than the 40–49 group (MD = 8.77, t = 3.03, p = .034). Similarly, there were significant group differences in years spent in education. Participants in the 30–39 age group spent longer in education compared to the 50–59 age group (MD = 3.09, t = 3.06, p = .031). There were also differences between the groups in technological familiarity. The 20–29 age group had significantly higher TFQ scores than the 60–69 (MD = 11.21, t = 5.05, p < .001) and 70–79 (MD = 16.22, t = 7.40, p < .001) age groups. The 30–39 group also had significantly higher scores compared the 60–69 (MD = 10.91, t = 4.78, p < .001) and 70–79 (MD = 15.91, t = 7.05, p < .001) age groups. Similarly, the 40–49 age group scored higher than the 60–69 (MD = 7.00, t = 3.03, p = .034) and 70–79 (MD = 12.01, t = 5.26, p < .001) age groups. Finally, the 50–59 age group had higher TFQ scores compared to the 70–79 age group (MD = 10.39, t = 4.60, p < .001). With regards to ethnicity, 78% of the sample were from White (n = 106), 7% were from Black (n = 9), 9% were from an Asian (n = 13) background, while 5% were Mixed race (n = 7), and one participant was Bengali (1%).

Distributions and descriptive statistics for main VStore outcomes

Table 2 below summarise the descriptive statistics for main VStore outcomes stratified by age group. We also present the frequency and density distributions of these outcomes (Figure 5). The Shapiro–Wilk test of normality indicated that all main outcome variables deviated from the normal distribution.

Group differences in overall VStore performance between age groups

There was a main effect of age on VStore Total Time, $X^2(5)$ = 37.71, p < .001; with a large effect size (r = 0.25, 95%CI = 0.15-0.42)—Figure 6. Benjamini-Hochberg corrected Dunn test revealed that the 70–79 age group (median = 932.85, IQR = 284.96) was significantly slower to complete VStore than the 20–29 (median = 543.45, IQR = 189.61), p < .001; 30–39 30–39 (median = 601.90, IQR = 189.61), p < .001; and 40–49 (median = 700.40, IQR = 221.16), p = .001, groups. In addition, the 60–59 age group (median = 748.79, IQR = 218.03), p = .002; and 50–59 age group (median = 810.43, IQR = 274.36), p = .002, groups were also significantly slower than the 20–29 age group. Adjusting for IQ and technological familiarity did not significantly alter results—all

TABLE 1 Study 1 demographic information.

	Age group						Statistic
	20–30	30–39	40–49	50–59	60–69	70–79	
N							
Count	26	22	23	23	22	23	NA
Percent	19%	16%	17%	17%	16%	17%	
Age							
Mean	23.81	32.87	45.05	53.70	63.95	73.65	NA
SD	2.37	2.26	2.90	2.85	2.65	3.02	
Gender							
Count	14 F	11 F	11 F	11 F	12 F	12 F	$X^2 = 0.40$, $p = .995$
Percent	54%	50%	48%	48%	55%	52%	
IQ							
Mean	117.19	120.96	115.82	120.39	124.59	126.26	F = 4.09, $p = .002$
SD	10.75	7.13	14.64	7.66	7.25	7.94	
Edu (years)							
Mean	17.65	19.26	18.05	16.17	16.73	17.00	F = 2.36, $p = .044$
SD	2.12	3.05	3.34	3.23	3.37	4.94	
Tech Fam							
Mean	47.35	47.04	43.14	41.52	36.14	31.13	F = 16.18, $p < .001$
SD	6.46	6.05	7.88	9.39	7.70	8.18	

Notes. N, sample size; IQ, intelligence quotient; Edu (years); years spent in education; Tech Fam; technological familiarity.
SD, standard deviation; F, female; NA, not applicable; X^2, Chi-square; F, F-value; p, p-value.

group differences remained significant expect for the difference between the 20–29 and 60–69 age groups ($p = .089$).

Study 2: Virtual reality cognitive assessment in healthy younger and older adults

Materials and methods

Participants

Forty healthy participants between 20 and 30 ($n = 20$), and 60 and 70 ($n = 20$) years of age took part in the study. Opportunistic sampling was used to recruit volunteers from the South London area through advertisements placed in community businesses, and on social media and local community websites. An equal number of males and females ($n = 10$) were recruited into both age groups. Analogous to Study 1, criteria for inclusion required the absence of a current Axis I disorder (DSM-5), alcohol and/or substance use disorder, neurological illnesses, a history of severe motion sickness, a history of head injury resulting in the loss of consciousness, mobility difficulties, and pregnancy.

Measures

Measures used in Study 2 are almost identical to measures used in Study 1 (see Section 2.1.2). VStore was completed with a marginally altered shopping list. Lemon Curd was replaced with Raspberry Jam to remove cultural bias, as many non-British participants were not familiar with the preserve in Study 1. Furthermore, Potato was replaced with Baked Beans to ensure that participants need to explore all aisles in the shop.

The TFQ and WASI-I was completed as described in Study 1. The internal consistency of the TFQ, while not as high as in Study 1, was still adequate (Cronbach's $\alpha = 0.83$, 95%CI = 0.82–0.89).

The main modification between Study 1 and 2 is the inclusion of the Virtual Reality Symptom Questionnaire (VRSQ) (Ames et al., 2005). The VRSQ measures adverse symptoms that may be induced by immersive VR such as general discomfort, dizziness, eyestrain, and headache. Symptoms are evaluated on a 7-point likert scale indicating whether the participant experienced no (0), mild (1–2), moderate (3–4), or severe (5–6) adverse effects due to VR exposure. The maximum score obtainable on the measure is 78. The internal consistency of the questionnaire was high at pre- (Cronbach's $\alpha = 0.93$, 95%CI = 0.90–0.96), and good at post-VR (Cronbach's $\alpha = 0.87$, 95%CI = 0.81–0.92) exposure.

Procedures

Eligible participants were invited for a single four-hour-long study visit following a telephone screen. Informed consent was obtained, in addition to demographics, and mental and physical health history. Electroencephalogram (EEG) data acquisition

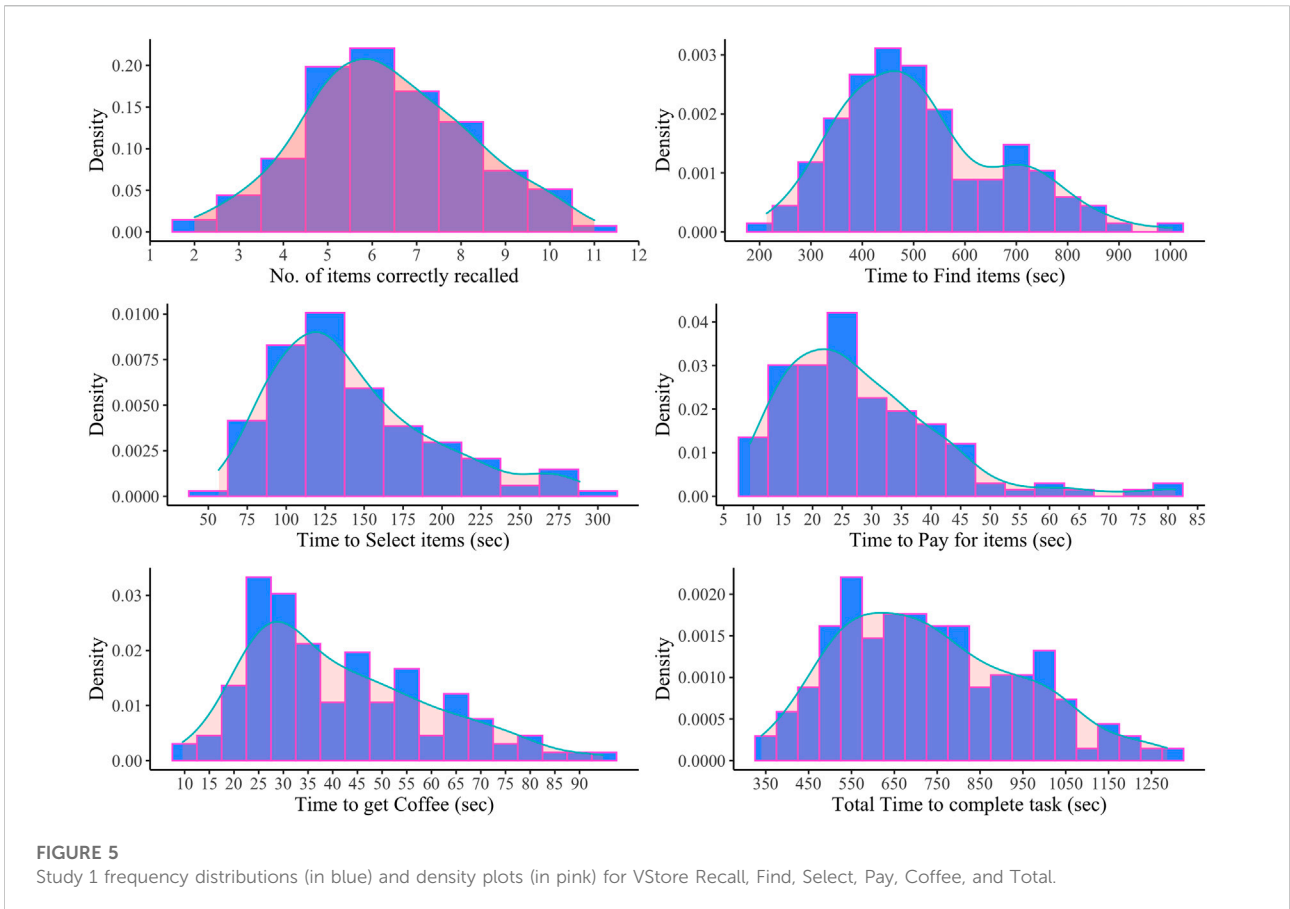TABLE 2 Study 1 descriptive statistics including means, standard deviations, and range for VStore outcomes.

	Age group					
	20–30	**30–39**	**40–49**	**50–59**	**60–69**	**70–79**
Recall						
Mean (count)	6.65	6.52	6.76	6.09	6.55	5.36
Standard deviation	1.90	1.95	1.76	1.80	1.84	1.73
Minimum	10	10	11	10	10	9
Maximum	2	3	4	3	3	2
Find	10	10	11	10	10	9
Mean (seconds)	419.37	462.4	475.04	536	546.15	655.95
Standard deviation	134.83	147.85	115.75	140.48	122.54	168.80
Minimum	213.48	248.49	287.14	304.64	375.79	403.37
Maximum	785.72	790.52	739.95	791.30	856.21	1004.00
Select						
Mean (seconds)	118.29	126.25	125	168.45	155.93	166.85
Standard deviation	43.64	52.57	46.03	53.34	45.23	48.20
Minimum	76.44	56.70	66.97	105.22	99.29	84.66
Maximum	269.95	255.29	276.30	266.70	288.29	286.26
Pay						
Mean (seconds)	26.25	22.67	25.23	25.97	29.51	41.44
Standard dev.	15.72	8.38	10.34	9.88	10.35	17.48
Minimum	10.71	9.44	11.68	11.56	11.93	12.71
Maximum	81.00	43.87	45.78	43.53	49.90	81.24
Coffee (seconds)						
Mean	31.8	34.01	32.63	50.3	48.92	50.46
Standard deviation	14.97	14.69	10.12	18.06	18.43	18.77
Minimum	12.54	9.58	20.93	26.37	20.53	29.72
Maximum	78.64	63.83	60.94	94.77	83.12	90.77
Total (seconds)						
Mean	595.03	668.34	671.16	782.42	780.21	932.02
Standard deviation	175.94	226.91	149.09	183.22	156.95	198.12
Minimum	344.14	339.66	419.81	452.69	544.94	577.17
Maximum	1046.30	1286.80	947.90	1082.89	1053.43	1262.86

followed to test how VStore performance relates to EEG markers of cognitive decline—these results will be reported elsewhere (Patchitt et al., 2022). Participants completed the TFQ during cap fitting, after which they were given a 30-min break to wash their hair and take lunch. After the break, participants completed cognitive assessments including VStore and the Cogstate computerised battery. Finally, the WASI-I was completed. Participants were compensated for their time and reimbursed for travel and sustenance. Ethics Approval was granted by the Psychiatry, Nursing and Midwifery Research Ethics Committee, King's College London (HR-18/19-11868).

Analysis

Analogous to Study 1, feasibility and acceptability was evaluated by calculating the overall completion rate for VStore, expressed as a percentage. Demographic characteristics for both groups were examined using means, SDs, counts, and percentages. Group differences in gender and eyesight ratio were tested using Chi-square. Group differences in IQ, education years, and TFQ total score was probed using independent samples t-test. Descriptive statistics for main VStore outcomes were calculated, and sample distributions visually evaluated using histograms stratified by group and density plots for the overall sample. Prior to analysis, extreme outliers (x −/+ 2.5 SD) were removed. Two participants, both in the 60–70 age group, with outlier values on more than one VStore outcome were eliminated, and one single datapoint was removed from the VStore Pay outcome variable. Deviations from the normal distribution were evaluated statistically using the Shapiro-Wilk test. As an exploratory aim, differences in

FIGURE 5
Study 1 frequency distributions (in blue) and density plots (in pink) for VStore Recall, Find, Select, Pay, Coffee, and Total.

overall VStore performance (i.e., Total Time) was assessed using independent samples *t*-test and Cohen's d for effect size. *t*-test was selected based on the results of the Shapiro-Wilk test. An ANCOVA was also performed to account for group differences in IQ and technological familiarity. Finally, VStore tolerability was probed by assessing pre- and post-VR symptoms, as measured by the VRSQ, using the Wilcoxon matched pairs signed rank test.

Results

VStore completion rate

All participants successfully completed VStore. However, two participants from the 60–70 age group had to be subsequently removed due to outliers. Taking these exclusions into account, VStore completion rate was 95%; otherwise, the completion rate was 100%.

Sample characteristics

Demographic information is presented in Table 3. The sample was not ethnically diverse—79% of participants were white (*n* = 30), 13% were Asian (*n* = 5), and 8% were Mixed Race

(*n* = 3). There were some notable group differences; the 60–70 age group had a higher IQ score on average, while the 20–30 age group achieved a higher technological familiarity total score.

Distributions and descriptive statistics for main VStore outcomes

Table 4 below summarises the descriptive statistics for main VStore outcomes. In addition, we present the frequency distributions of these variables stratified by group status as well as density distributions for the complete sample in Figure 7. The Shapiro–Wilk test is a test of normality indicated that VStore Find deviated significantly from the normal distribution.

Group differences in overall VStore performance between age groups

There was a significant difference in VStore Total Time between age groups ($t_{(32.24)=}$ 7.76 $p < .001$), with a large effect size (Cohen's $d = 2.54$, 95%CI = 1.93–3.65). Participants aged 20–30 completed VStore in 506.97 s (SD = 92.19) on average compared to participants aged 60–70 who took 774.59 s (SD = 117.32) to complete the task (Figure 8).

FIGURE 6
Study 1 group differences in VStore Total Time performance.
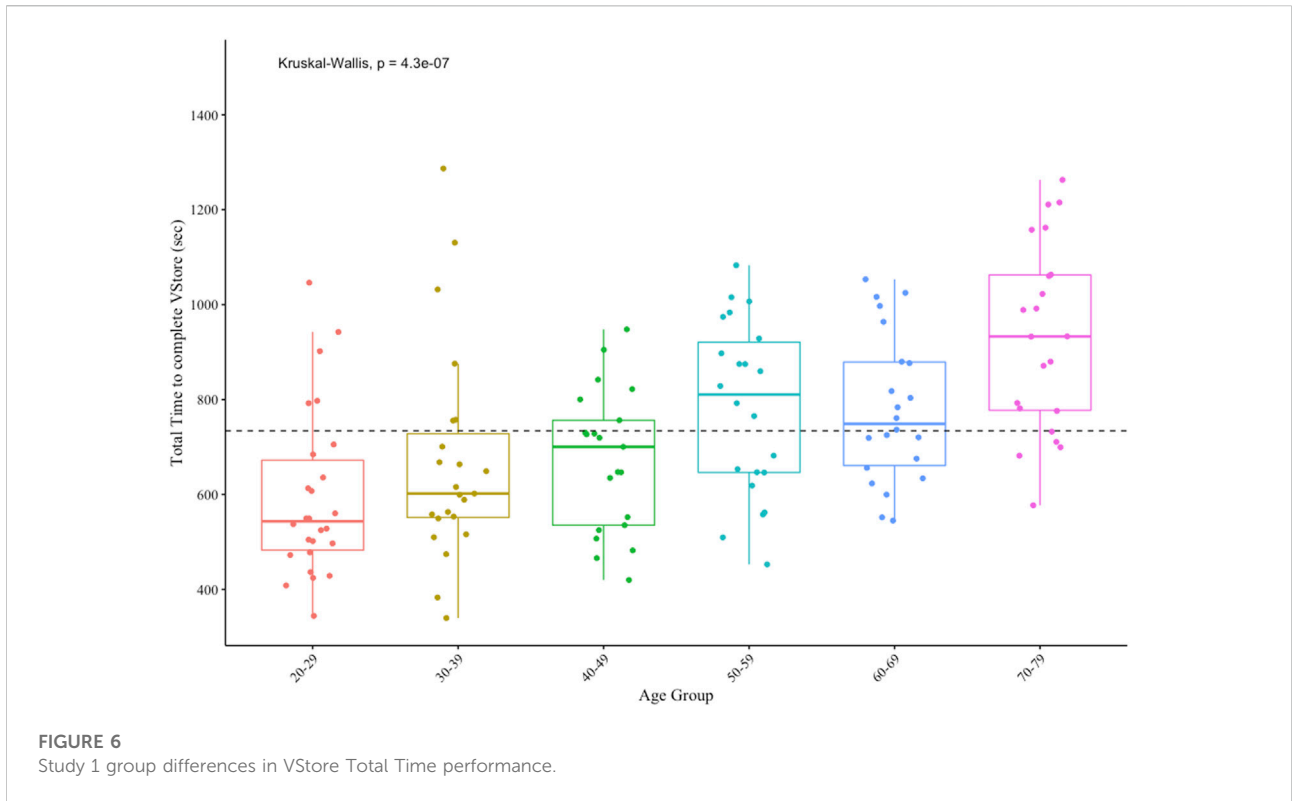
Adjusting for IQ and technological familiarity did not alter results.

VStore acceptability and adverse effects

VR exposure did not cause any adverse effects. In fact, participants reported fewer symptoms post-VStore. The median VRSQ total score was 1 pre-VStore (IQR = 4) and 0 post-VStore (IQR = 1); V = 308, p = .001. The median was 0 both pre- and post-VStore for dizziness, eyestrain, blurred vision, headache, and general discomfort.

Study 3: Virtual reality cognitive assessment in patients with psychosis

Materials and methods

Participants

Twenty-eight male and female patients aged 18 to 60 with schizophrenia or a related psychotic-spectrum disorder took part in the study. Participants were recruited from outpatient services at the South London and Maudsley (SLaM) National Health

TABLE 3 Study 2 demographic information.

	20–30		60–70		Statistics
	Mean/Count	SD/%	Mean/Count	SD/%	
N	20	NA	18	NA	NA
Age	24.10	2.99	65.78	2.39	NA
Gender	10 F	50%	9 F	50%	ns
IQ	117.60	8.21	125.28	7.13	$t_{(36.5)}$ = 3.06, p = .004
Edu Years	16.05	2.48	16.22	3.54	ns
TFQ	45.85	6.24	39.06	9.03	$t_{(31.5)}$ = 2.67, p = .012
Eyesight	6 corrected	30%	6 corrected	33%	ns

Notes. SD, standard deviation; %, percentage; NA, not applicable; F, female; ns; not significant; t, t-value; p, p-value.
N, sample size; IQ, intelligence quotient; Edu Years, years spent in education; TFQ, technological familiarity questionnaire.

TABLE 4 Study 2 descriptive statistics for VStore outcomes in seconds stratified by group.

	20–30			60–70		
	Mean	**SD**	**Range**	**Mean**	**SD**	**Range**
Recall	6.25	2.24	2–9	5.50	1.95	3–10
Find	349.54	58.93	263.13–463.62	537.88	102.09	390.60–754.15
Select	107.45	36.25	57.35–223.96	169.00	44.34	87.37–262.13
Pay	18.29	7.41	9.58–32.348	23.03	8.02	8.00–38.19
Coffee	30.39	12.63	15.21–65.29	45.36	13.18	24.90–70.53
Total	506.97	92.19	376.51–709.80	774.59	117.32	583.99–991.50

Notes. SD, standard deviation.

Service Foundation Trust. Eligible participants had to be stable on their antipsychotic medication for at least one month, and able to provide informed consent. Patients with an Axis I disorder other than psychosis (DSM-5), moderate to severe alcohol and/or substance use disorder, a history of severe motion sickness, a neurological illness, IQ below 70, mobility issues, and who are pregnant were excluded.

Measures

Mini international neuropsychiatric interview 7.0.2 for DSM-5 (M.I.N.I.) (Sheehan et al., 1998)
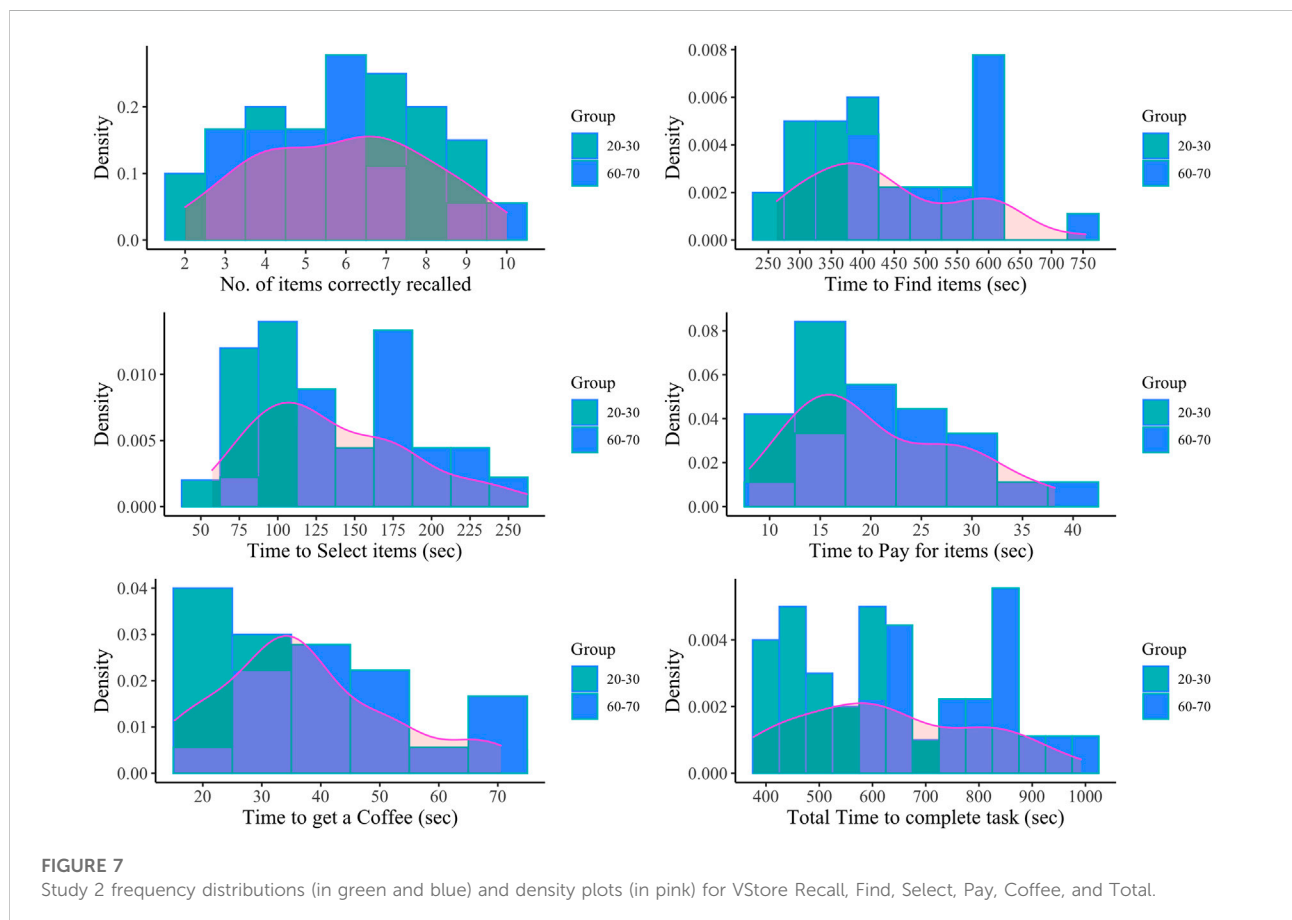
The M.I.N.I. is a short structured diagnostic interview including yes/no questions used to probe the presence (and absence) of Axis-I psychiatric disorders. In this study, the MINI was used to establish eligibility.

Wechsler abbreviated scale of intelligence—1st edition

Consistent with Study 1 and 2, the WASI-I was completed to establish the age relative IQ score of participants, using the matrix and vocabulary tests.

Extended technological familiarity questionnaire

We developed an extended version of the Technological Familiarity Questionnaire used in the previous studies (Porffy et al., 2022). The eTFQ consists of three subscales measuring knowledge, proficiency, and attitudes towards seven popular devices (computer, smartphone, tablet, smartwatch, smart TV, console games, and VR), and technology in general (Supplementary Material S4). Higher scores indicate better familiarity with technology. The internal consistency of the
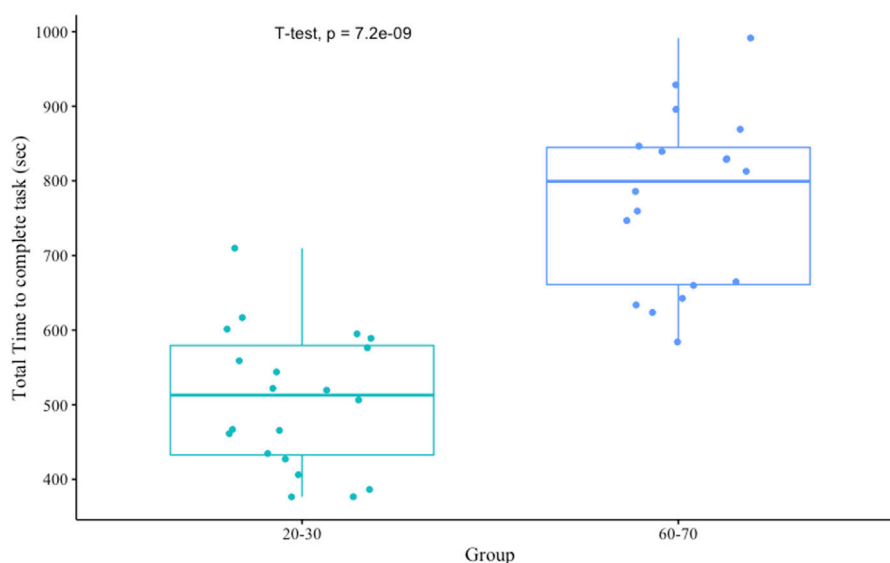

FIGURE 7
Study 2 frequency distributions (in green and blue) and density plots (in pink) for VStore Recall, Find, Select, Pay, Coffee, and Total.

FIGURE 8
Study 2 Group Differences in VStore Total Time performance.

eTFQ was excellent (Cronbach's α = 0.94, 95%CI = 0.92–0.97). The questionnaire is currently undergoing validation.

Technological familiarity questionnaire—12-item

Twelve items of the eTFQ form the TFQ except for last question assessing "ability to use technology", which has no equivalent in the extended version. To provide an indication of patients' technological familiarity relative to healthy volunteers in studies 1 and 2, we calculated the TFQ-12 score. The internal consistency of the TFQ-12 was adequate (Cronbach's α = 0.88, 95%CI = 0.81–0.94).

VStore

VStore was completed as described in the introduction. The exact same shopping list was used as in Study 2.

Virtual reality symptom questionnaire

The VRSQ was again used to assess adverse effects induced by VR. The internal consistency of the questionnaire was good both pre- (Cronbach's α = 0.87, 95%CI = 0.80–0.94) and post-VStore (Cronbach's α = 0.88, 95%CI = 0.82–0.93) administration.

Positive and negative syndrome scale (Kay et al., 1987)

The PANSS semi-structured interview was used to assess severity of psychotic symptoms on a 30-item rating scale, probing positive, negative, and general symptoms.

Procedures

Potential participants were pre-screened over the phone. Those appearing to be eligible were invited for an in-person screening visit during which the MINI 7.0.2, WASI-I, and eTFQ were completed. Those meeting the inclusion criteria, returned to complete VStore, the VRSQ, and the PANSS, alongside other measures including the MATRICS Consensus Cognitive Battery and University of California San Diego Performance-Based Skills Assessment, and additional functional capacity measures. Participants were compensated for their time and reimbursed for travel expenses and sustenance. Ethical approval was granted by the Health Research Authority and London Bromley Research Ethics Committee (19/LO/0792).

Analysis

Corresponding to Study 1 and 2, feasibility and acceptability were assessed by the overall completion rate of VStore. Demographic characteristics, including age, gender, IQ, years spent in education, eyesight, diagnosis, symptom profile, duration of illness, chlorpromazine daily oral dose equivalent (Leucht et al., 2016), depot and antidepressant use were tested using means, SDs, ranges, and percentages. Descriptive statistics for main VStore outcome variables were calculated, and sample distributions visually depicted using histograms and density plots. Three outlier values from two participants were removed from main VStore outcomes (one from Select, one from Pay, and one from Total Time) as they were more than 2.5 SDs away more from the sample mean. Given the relatively small sample size, no complete patient dataset was removed to preserve statistical power. Deviations from the normal distribution was tested statistically using the Shapiro-Wilk test. Tolerability was assessed pre- and post-VR exposure, as

measured by the VRSQ, using the Wilcoxon matched pairs signed rank test. The Wilcoxon test was selected as the data was not normally distributed. Finally, patients from this study were compared to age (within 2 years) and gender matched healthy participants from Study 1 on VStore Total Time using the Wilcoxon test and effect size (r). Three patients and their control pairs were excluded from this analysis due to outlier values on the VStore outcome (x −/+ 2.5 SD).

Results

VStore completion rate

Out of 28 participants, one was excluded due to researcher's error during the set-up of VStore—11 instead of 12 items were added to the shopping list. If this mistake is considered, the completion of VStore was 96.43%. Otherwise, the completion rate was 100%.

Sample characteristics

Demographic information is presented in Table 5. In contrast with Study 1 and 2, the sample was ethnically diverse—only 29% of participants were from white (n = 8), 63% were black (n = 17), and 4% were from Asian (n = 1), and 4% were from Mixed (n = 1) origin. Patients had a lower IQ score and spent fewer years in education compared to participants in Study 1 and 2.

Distributions and descriptive statistics for main VStore outcomes

Table 6 below summarise the descriptive statistics for main VStore outcomes. In addition, we present the frequency and density distributions of these in Figure 9. The Shapiro–Wilk test is a test of normality indicated that VStore Recall, Select, and Coffee deviated significantly from the normal distribution.

Group differences in overall VStore performance between patients and controls

There was a significant difference in VStore Total Time between patients in Study 3 and age and gender matched healthy participants in Study 1 (W = 456, p < .001), with a large effect size (r = 0.5, 95%CI = 0.25–0.70). Patients completed VStore in 970.99 s (IQR = 564.29) on average compared to healthy controls who took 608.83 s (IQR = 244.26).

VStore acceptability and adverse effects

VR exposure did not cause any adverse effects. Patients reported fewer symptoms following VStore administration, with a median VRSQ total score of 5 pre- (IQR = 9.5) and 2 post-VStore (IQR = 5.5); V = 110, p = .117. Similar to Study 2, the median for dizziness, eyestrain, blurred vision, headache, and general discomfort was 0 both pre- and post-VStore administration.

TABLE 5 Study 3 demographic information.

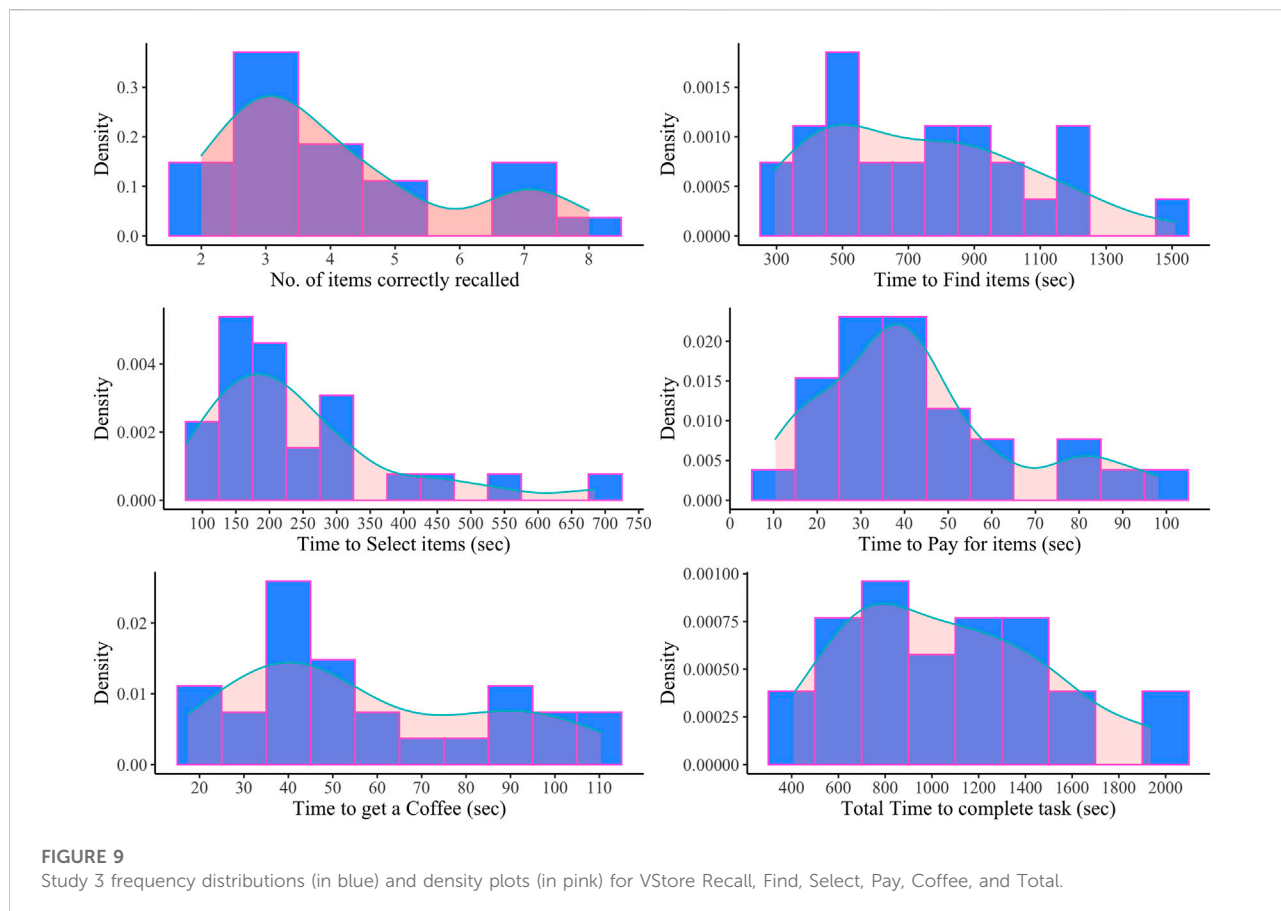|	Mean/Count	SD/%	Range
Age	40.63	10.93	20–60
Gender	6 Female	22%	NA
Intelligence Quotient	97.22	14.21	77–124
Education (Years)	14.52	2.86	11–23
eTFQ	190.20	65.69	83–331
TFQ-12	29.48	11.23	13–50
Eyesight	13 Corrected	48%	NA
Diagnosis			
Schizophrenia	15	56%	NA
Schizoaffective Disorder	6	22%	NA
Unspecified Psychosis	6	22%	NA
PANSS			
Positive	14.26	4.75	7–24
Negative	12.63	4.01	7–19
General	27.22	4.75	16–42
Total	54.11	13.14	31–85
Duration of Illness	12.96	10.83	1–35
Chlorpromazine Equivalent	236.21	157.51	90–600
Depot Injection	9 Patients	33%	NA
Antidepressant Medication	3 Patients	11%	NA

Notes. eTFQ, Extended Technological Familiarity Questionnaire; TFQ-12, Technological Familiarity Questionnaire 12-item; PANSS, Positive and Negative Syndrome Scale; SD, standard deviation; %, percentage; NA, not applicable.

TABLE 6 Study 3 descriptive statistics for VStore outcomes.

	Mean	Standard Deviation	Range
Recall	4.04	1.76	2–8
Find	744.99	318.61	294.21–1509.56
Select	246.60	142.34	75.9–685.23
Pay	43.16	22.81	10.41–98.18
Coffee	57.21	28.37	17.45–110.35
Total	1062.93	422.44	408.83–1936.35

General discussion

Our primary aim was to establish the feasibility and acceptability of administering VStore in three separate samples. Feasibility was assessed by overall completion rate. High rates were achieved across all three studies. In Study 1, including healthy volunteers between the ages 20 and 79, the completion rate was 99.30%. Only one participant failed to complete VStore in the 70–79 age group due to fatigue. The completion rate in Study 2, including two groups of healthy participants aged 20–30 and 60–70, was 100%. In Study 3, including patients with psychosis, the completion rate was

FIGURE 9
Study 3 frequency distributions (in blue) and density plots (in pink) for VStore Recall, Find, Select, Pay, Coffee, and Total.

again 100%. The three studies included 210 participants in total, and the overall completion rate across the three samples was 99.95%. These results indicate that VStore is acceptable and feasible to administer to healthy adults across the age spectrum and provide preliminary evidence of feasibility in psychosis.

A comparable study to ours, evaluating the feasibility of administering a VR shopping assessment to patients following stroke, found that the within-sample completion rate was lower in the HMD condition (84%) compared to the computer screen display condition (93%) (Spreij et al., 2020). Similar results were reported in the control group—91% in the HMD condition *vs.* 100% computer screen display condition. Indeed, while both patient and controls reported enhanced feelings of engagement, flow, and presence with HMD, they also complained of more adverse effects. Overall, they reported no preference for either interface, prompting the authors to conclude that both are feasible to employ. This is in line with our findings showing that a fully immersive VR shopping task is feasible to administer to healthy individuals. On the other

hand, we did not find any evidence for HMD inducing unwanted side effects in our samples.

In studies 2 and 3, participants completed a questionnaire pre- and post-VStore immersion evaluating symptoms related to cybersickness. VStore did not induce any adverse effects commonly associated with HMD such as general discomfort, dizziness, eyestrain, or blurred vision. Patients tended to score higher on these items both pre- and post-VStore immersion. In Study 3, the median total score was five pre-VStore and two post-VStore; while in study 2, it was one pre-VStore and zero post-VStore, indicating that patients have a higher baseline level of discomfort. These results together with high completion rates suggest that VStore is extremely well-tolerated. This could be explained by the technological features of the HTC Vive, PC set-up, and mindful software development. Evidence suggests that old generation HMD induces more cybersickness compared to new generation systems, and the least adverse effects are observed with commercial kits (as compared to development kits), such as the one used in this study (Kourtesis et al., 2019). Other important factors include: 1) good quality display (OLED >

LCD), 2) adequate field of view (FOV ≥110°), 3) adequate resolution per eye (>960 × 1080 sub-pixels), and 4) high refresh rate (>75 Hz) (Kourtesis et al., 2019)—all of which were met or exceeded. In terms of computer hardware, we used a powerful PC with a high-quality gaming video card to meet the computing power requirements of both the headset and software (Anthes et al., 2016), and thus ensure the smooth running of the assessment. Finally, VStore was purposefully designed to minimise unwanted side effects through employing instant teleportation, which is thought to be the most suitable method for movement in VR (Bozgeyikli et al., 2016); and requiring the users to carry out ergonomic interactions in the environment, such as moving their heads to look around or picking up objects (Kourtesis et al., 2019).

We also examined the distribution of VStore outcomes through descriptive statistics, histograms, and density plots. The number of items correctly recalled in VStore ranged from 2 to 11 in Study 1, 2–10 in Study 2, and 2-8 in Study 3. Given that no participant recalled the maximum number of items (12), it is unlikely that this outcome is susceptible to ceiling effect. VStore Recall deviated from the normal distribution in two of the studies; therefore, when analysed, non-parametric procedures should be considered. VStore Find was the only variable that significantly deviated from the normal distribution in all three studies. This is not unexpected as reaction time data is often skewed, and thus should be log transformed prior to analysis (Van Der Linden, 2006), and/or non-parametric analysis methods should be considered. The fastest participant (213 s) on this outcome was in the youngest age group and the slowest performance (1004 s) was found in the oldest age group in Study 1, as expected. The same trend was observed in Study 2, where VStore Find times ranged from 263 to 754 s. Patients in Study 3 were notably slower with a mean completion time of 745 s (range = 294–1509). Differences in performance between age groups and patients vs. healthy participants, suggest that VStore Find may be able to differentiate between these groups. VStore Select was positively skewed in all three studies and thus should be log transformed. Again, patients completed this step slower compared to healthy volunteers in studies 1 and 2. In Study 1, the fastest performance on VStore Select (56 s) was in the 30–39 group, and the slowest (288 s) was in the 60–69 group, indicating that this outcome may not be as helpful in distinguishing between age groups as VStore Find. VStore Pay also showed a positive skew and patients took longer on average than healthy participants; however, performance in the 70–90 group in Study 1 was comparable to that of patients in Study 3. A similar trend was observed in VStore Coffee with regards to distribution, but average performance of patients was not markedly different from healthy individuals. Finally, overall VStore performance measured by Total Time spent completing the task only

significantly deviated from the normal distribution in Study 1 but was positively skewed in all studies, and thus log transform should be considered. With regards to distinguishing between study groups, there is a clear difference in VStore Total Time between patients and healthy participants and younger and older age groups, suggesting that comparably to VStore Find, this outcome could potentially be used to differentiate between age groups and healthy and clinical populations.

Previous research into VR neuropsychological assessments found that these types of measures are sensitive to age and younger participants tend to complete them faster (Davison et al., 2018; Oliveira et al., 2018). Our results are in line with these findings. In Study 1, participants aged 70–79 were significantly slower to complete VStore than participants aged 20–49, and participants aged 50–69 were slower than the youngest age group (20–29). We replicated these findings in Study 2, where participants aged 20–30 completed VStore significantly faster than participants aged 60–70. In Study 1, the cut-off for noticeable differences in performance was found to be at the 50-years mark. While there is evidence suggesting that cognitive function primarily reflects person-specific factors and decline may start in healthy educated adults when they are in their late 20s and early 30s (Wilson et al., 2002; Salthouse, 2009), it is widely thought that marked decline in cognition presents when adults reach their 50s (Albert, 1981; Rönnlund et al., 2005)—this is supported by our findings. However, there may be an alternative explanation. Technological familiarity decreases with age (Olson et al., 2011), which can have a significant impact on the outcomes of cognitive assessments embedded in VR and may result in the underestimation cognitive abilities in this population; hence, the observed difference in VStore performance.

It is important to note that there are a number of limitations linked to sampling. Studies 1 and 2 are likely not representative of the general population. Participants are highly educated and have a high IQ on average; and the use of the abbreviated IQ measure relying on two domains only—verbal ability and matrix reasoning—may generate inflated scores (Axelrod, 2002). Our exclusion criteria also preclude the generalisation of these findings to the wider population. For example, we excluded individuals with severe motion sensitivity and pregnant women, as they are more likely to experience motion sickness. In addition, we excluded people with mobility issues. These factors should be considered when interpreting these findings. On the other hand, studies 1 and 2 were largely ethnically representative of the population of England and Wales based on the 2011 Census data (Office for National Statistics, 2015). In Study 3, patients from Black origin were overrepresented compared to White, Asian, and Mixed ethnic backgrounds, which may be explained by the elevated risk of this ethnic group being diagnosed with a psychotic disorder (Halvorsrud et al., 2019).

Beside the issues around sampling, there are further limitations to consider. Potential adverse effects were not measured in Study 1. It is important to note here that no participant stopped VStore due to cybersickness. Only one participant, aged 79, had to discontinue the assessment due to exhaustion, after being immersed for 37 min. No cybersickness was reported in the other samples. Sense of presence was not measured in any of the studies. This is a major limitation considering that movement was mainly executed through teleportation which may affect the sense presence and ecological validity (Weech et al., 2019). Another important limitation is the lack of quantitative and qualitative feedback from participants on their experience of VStore. While we did not collect formal responses, in studies 2 and 3, some participants left written feedback on the post-VRSQ questionnaire under "other symptoms and feelings". These were all positive. Some examples include: "joy", "sense of achievement", "I feel refreshed like I've accomplished something I could not do before", and "intrigued". Finally, there were some group differences in IQ, years spent in education, and technological familiarity between age groups in studies 1 and 2, which may contribute to the observed differences.

Based on our experience of conducting these studies and the available literature, we would like to make some suggestions for the development and testing of VR assessments in mental health. As recommended by VR-CORE, content development—akin to video game development—should involve end-users in early software design and testing. Careful considerations should be given to the minimisation of unwanted effects. For example, we found that teleportation in conjunction with physical movement works well and does not induce cybersickness. Physical abilities and the mental health of patients must also be considered from the early stages—we found, for example, that patients with psychosis have higher levels of baseline discomfort compared to healthy individuals. The second stage of VR assessment validation should focus on the collection of feasibility, acceptability, and tolerability data, and provide an initial indication of its potential uses. At this stage, both quantitative and qualitative data on user experience should be collected, and if needed, further tweaks to the software should be carried out. The final stage of the validation should be conducted in large, representative samples with the aim to evaluate clinical utility by comparing clinical populations to controls and existing gold-standard measures to the novel VR assessment. A longitudinal design also enables the evaluation of test-retest reliability, practice effect, and the instrument's sensitivity to changes in performance.

In conclusion, we demonstrated that VStore is a promising neuropsychological assessment that is well-tolerated and feasible to administer to healthy adults and patients with psychosis. We also found evidence that VStore can differentiate between age groups, suggesting that it may be suitable to study age-related cognitive decline. However, this requires further research to confirm. Future studies should focus on comparing VStore to standard cognitive and functional capacity measures both in healthy individuals and clinical populations affected by cognitive decline and related functional difficulties. Some of this work has already been carried out and published demonstrating that VStore engages the same cognitive domains as measured by the Cogstate computerised cognitive battery in healthy volunteers (Porffy et al., 2022), and more studies are currently underway in clinical populations. Nonetheless, these findings would have to be replicated in large independent samples and further work is needed to establish VStore's ecological validity.

Data availability statement

The datasets presented in this article are not readily available due to commercial restrictions. VStore is a propriety software. Requests to access the datasets should be directed to LP.

Ethics statement

The studies involving human participants were reviewed and approved by Study 1: Psychiatry, Nursing and Midwifery Research Ethics Committee, Kings College London (LRS-16/17-4540). Study 2: Psychiatry, Nursing and Midwifery Research Ethics Committee, Kings College London (HR-18/19-11868). Study 3: Health Research Authority and London Bromley Research Ethics Committee (19/LO/0792). The patients/participants provided their written informed consent to participate in this study.

Author contributions

LP: Study design, data collection and analysis, written manuscript. MM: Advised on data analysis, manuscript preparation and review. EM: Software development, study design, contribution to manuscript. SS: Study design, manuscript preparation and review.

Funding

Acknowledgments

Conflict of interest

SS and EM created VStore, with contributions from LP and technical development from VitaeVR Ltd. King's College London has licensed its rights in VStore to VitaeVR Ltd. LP, and SS are entitled to a share of any revenues King's College London may receive from commercialisation of VStore by VitaeVR Ltd.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2022.875197/full#supplementary-material

References

Albert, M. S. (1981). Geriatric neuropsychology. *J. Consult. Clin. Psychol.* 49, 835–850. doi:10.1037/0022-006x.49.6.835

American Psychiatric Publishing (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. 5th ed. Washington, DC: American Psychiatric Publishing.

Ames, S. L., Wolffsohn, J. S., and Mcbrien, N. A. (2005). The development of a symptom questionnaire for assessing virtual reality viewing using a head-mounted display. *Optometry Vis. Sci.* 82, 168–176. doi:10.1097/01.OPX.0000156307.95086.6

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

Atkins, A. S., Khan, A., Ulshen, D., Vaughan, A., Balentin, D., Dickerson, H., et al. (2018). Assessment of instrumental activities of daily living in older adults with subjective cognitive decline using the virtual reality functional capacity assessment tool (VRFCAT). *J. Prev. Alzheimers Dis.* 5, 216–234. doi:10.14283/jpad.2018.28

Axelrod, B. N. (2002). Validity of the Wechsler Abbreviated Scale of Intelligence and other very short forms of estimating intellectual functioning. *Assessment* 9, 17–23. doi:10.1177/1073191102009001003

Bermudez-Contreras, E., Clark, B. J., and Wilber, A. (2020). The neuroscience of spatial navigation and the relationship to artificial intelligence. *Front. Comput. Neurosci.* 14, 63–16. doi:10.3389/fncom.2020.00063

Birckhead, B., Khalil, C., Liu, X., Conovitz, S., Rizzo, A., Danovitch, I., et al. (2019). Recommendations for methodology of virtual reality clinical trials in health Care by an international working group: Iterative study. *JMIR Ment. Health* 6, e11973. doi:10.2196/11973

Boletsis, C., and Cedergren, J. E. (2019). VR locomotion in the new era of virtual reality: An empirical comparison of prevalent techniques. *Adv. Human-Computer Interact.* 2019, 7420781. doi:10.1155/2019/7420781

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

Buchanan, R. W., Davis, M., Goff, D., Green, M. F., Keefe, R. S. E., Leon, A. C., et al. (2005). A summary of the FDA-NIMH-MATRICS workshop on clinical trial design for neurocognitive drugs for schizophrenia. *Schizophr. Bull.* 31, 5–19. doi:10.1093/schbul/sbi020

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

Chersi, F., and Burgess, N. (2015). The cognitive architecture of spatial navigation: Hippocampal and striatal contributions. *Neuron* 88, 64–77. doi:10.1016/j.neuron.2015.09.021

Clus, D., Larsen, M. E., Lemey, C., and Berrouiguet, S. (2018). The use of virtual reality in patients with eating disorders: Systematic review. *J. Med. Internet Res.* 20, e157–9. doi:10.2196/jmir.7898

Coughlan, G., Laczó, J., Hort, J., Minihane, A. M., and Hornberger, M. (2018). Spatial navigation deficits — overlooked cognitive marker for preclinical Alzheimer disease? *Nat. Rev. Neurol.* 14, 496–506. doi:10.1038/s41582-018-0031-x

Davison, S. M. C., Deeprose, C., and Terbeck, S. (2018). A comparison of immersive virtual reality with traditional neuropsychological measures in the assessment of executive functions. *Acta Neuropsychiatr.* 30, 79–89. doi:10.1017/neu.2017.14

Eichenberg, C., and Wolters, C. (2012). *Virtual realities in the treatment of mental disorders: A review of the current state of research*. Editor C. Eichenberg (Rijeka, Croatia: InTech). doi:10.5772/50094

Fervaha, G., Zakzanis, K. K., Foussias, G., Graff-Guerrero, A., Agid, O., and Remington, G. (2014). Motivational deficits and cognitive test performance in schizophrenia. *JAMA Psychiatry* 71, 1058–1065. doi:10.1001/jamapsychiatry.2014.1105

Fornells-Ambrojo, M., Barker, C., Swapp, D., Slater, M., Antley, A., and Freeman, D. (2008). Virtual reality and persecutory delusions: Safety and feasibility. *Schizophrenia Res.* 104, 228–236. doi:10.1016/j.schres.2008.05.013

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., et al. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychol. Med.* 47, 2393–2400. doi:10.1017/S003329171700040X

Guna, J., Geršak, G., Humar, I., Krebl, M., Orel, M., Lu, H., et al. (2020). Virtual reality sickness and challenges behind different technology and content settings. *Mob. Netw. Appl.* 25, 1436–1445. doi:10.1007/s11036-019-01373-w

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., and Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature* 436, 801–806. doi:10.1038/nature03721

Halvorsrud, K., Nazroo, J., Otis, M., Brown Hajdukova, E., and Bhui, K. (2019). Ethnic inequalities in the incidence of diagnosis of severe mental illness in England: A systematic review and new meta-analyses for non-affective and affective psychoses. *Soc. Psychiatry Psychiatr. Epidemiol.* 54, 1311–1323. doi:10.1007/s00127-019-01758-y

Harvey, P. D., Khan, A., and Keefe, R. S. E. (2017). Using the positive and negative syndrome scale (PANSS) to define different domains of negative symptoms: Prediction of everyday functioning by impairments in emotional expression and emotional experience. *Innov. Clin. Neurosci.* 14, 18–22. Available at: https://pubmed.ncbi.nlm.nih.gov/29410933.

Heckers, S. (2001). Neuroimaging studies of the hippocampus in schizophrenia. *Hippocampus* 11, 520–528. doi:10.1002/hipo.1068

Herweg, N. A., and Kahana, M. J. (2018). Spatial representations in the human brain. *Front. Hum. Neurosci.* 12, 297. doi:10.3389/fnhum.2018.00297

Kay, S. R., Fiszbein, A., and Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* 13, 261–276. doi:10.1093/schbul/13.2.261

Keefe, R. S. E., Buchanan, R. W., Marder, S. R., Schooler, N. R., Dugar, A., Zivkov, M., et al. (2013). Clinical trials of potential cognitive-enhancing drugs in

schizophrenia: What have we learned so far? *Schizophr. Bull. Bp.* 39, 417–435. doi:10.1093/schbul/sbr153

Knopman, D. S., Jones, D. T., and Greicius, M. D. (2021). Failure to demonstrate efficacy of aducanumab: An analysis of the EMERGE and ENGAGE trials as reported by Biogen, December 2019. *Alzheimer's. &amp. Dement.* 17, 696–701. doi:10.1002/alz.12213

Kourtesis, P., Collina, S., Doumas, L. A. A., and MacPherson, S. E. (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

Leucht, S., Samara, M., Heres, S., and Davis, J. M. (2016). Dose equivalents for antipsychotic drugs: The DDD method: Table 1. *Schizophr. Bull. Bp.* 42, S90–S94. doi:10.1093/schbul/sbv167

Migoya-Borja, M., Delgado-Gómez, D., Carmona-Camacho, R., Porras-Segovia, A., López-Moriñigo, J. D., Sánchez-Alonso, M., et al. (2020). Feasibility of a virtual reality-based psychoeducational tool (VRight) for depressive patients. *Cyberpsychology, Behav. Soc. Netw.* 23, 246–252. doi:10.1089/cyber.2019.0497

Millan, M. J., Agid, Y., Brüne, M., Bullmore, E. T., Carter, C. S., Clayton, N. S., et al. (2012). Cognitive dysfunction in psychiatric disorders: Characteristics, causes and the quest for improved therapy. *Nat. Rev. Drug Discov.* 11, 141–168. doi:10.1038/nrd3628

Negut, A., Matu, S. A., Sava, F. A., and David, D. (2016). Virtual reality measures in neuropsychological assessment: A meta-analytic review. *Clin. Neuropsychol.* 30, 165–184. doi:10.1080/13854046.2016.1144793

Office for National Statistics (2015). 2011 Census analysis: Ethnicity and religion of the non-UK born population in England and Wales. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/2011censusanalysisethnicityandreligionofthenonukbornpopulationinenglandandwales/2015-06-18#:~:text=In%202011%2C%207.5%20million%20people,African%2FCaribbean%2FBlack%20British (Accessed January 5, 2022).

O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* 34, 171–175. doi:10.1016/0006-8993(71)90358-1

Oliveira, C. R., Filho, B. J. P. L., Esteves, C. S., Rossi, T., Nunes, D. S., Lima, M. M. B. M. P., et al. (2018). Neuropsychological assessment of older adults with virtual reality: Association of age, schooling, and general cognitive status. *Front. Psychol.* 9, 1085–1088. doi:10.3389/fpsyg.2018.01085

Olson, K. E., O'Brien, M. A., Rogers, W. A., and Charness, N. (2011). Diffusion of technology: Frequency of use for younger and older adults. *Ageing Int.* 36, 123–145. doi:10.1007/s12126-010-9077-9

Opriş, D., Pintea, S., García-Palacios, A., Botella, C., Szamosközi, Ş., and David, D. (2012). Virtual reality exposure therapy in anxiety disorders: A quantitative meta-analysis. *Depress. Anxiety* 29, 85–93. doi:10.1002/da.20910

Parsons, T. D., and Courtney, C. G. (2014). An initial validation of the virtual reality paced auditory serial addition test in a college sample. *J. Neurosci. Methods* 222, 15–23. doi:10.1016/j.jneumeth.2013.10.006

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

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

Patchitt, J., Porffy, L. A., Whomersley, G., Szentgyorgyi, T., Brett, J., Mouchlianitis, E., et al. (2022). *The relationship between virtual reality shopping task performance and EEG markers of ageing.* Available at: https://www.frontiersin.org/articles/10.3389/fnagi.2022.876832/full.

Plascencia-Villa, G., and Perry, G. (2020). "Chapter One - status and future directions of clinical trials in Alzheimer's disease," in *International review of neurobiology*. Editors G. Söderbom, R. Esterline, J. Oscarsson, and M. P. Mattson (Cambridge, MA: Elsevier Academic Press), 3–50. doi:10.1016/bs.irn.2020.03.022

Porffy, L. A., Mehta, M. A., Patchitt, J., Boussebaa, C., Brett, J., D'Oliveira, T., et al. (2022). A novel virtual reality assessment of functional cognition: Validation study. *J. Med. Internet Res.* 24, e27641–18. doi:10.2196/27641

Roettl, J., and Terlutter, R. (2018). The same video game in 2D, 3D or virtual reality – how does technology impact game evaluation and brand placements? *PLoS ONE* 13, e0200724–24. doi:10.1371/journal.pone.0200724

Rönnlund, M., Nyberg, L., Bäckman, L., and Nilsson, L.-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychol. aging* 20, 3–18. doi:10.1037/0882-7974.20.1.3

Rus-Calafell, M., Garety, P., Sason, E., Craig, T. J. K., and Valmaggia, L. R. (2017). Virtual reality in the assessment and treatment of psychosis: A systematic review of its utility, acceptability and effectiveness. *Psychol. Med.* 48, 362–391. doi:10.1017/S0033291717001945

Ruse, S. A., Harvey, P. D., Davis, V. G., Atkins, A. S., Fox, K. H., and Keefe, R. S. E. (2014). Virtual reality functional capacity assessment in schizophrenia: Preliminary data regarding feasibility and correlations with cognitive and functional capacity performance. *Schizophrenia Res. Cognition* 1, e21–e26. doi:10.1016/j.scog.2014.01.004

Sabbagh, M. N., Hendrix, S., and Harrison, J. E. (2019). FDA position statement "Early Alzheimer's disease: Developing drugs for treatment, Guidance for Industry. *Alzheimer's. &. Dementia Transl. Res. &. Clin. Interventions* 5, 13–19. doi:10.1016/j.trci.2018.11.004

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiol. Aging* 30, 507–514. doi:10.1016/j.neurobiolaging.2008.09.023

Segawa, T., Baudry, T., Bourla, A., Blanc, J. V., Peretti, C. S., Mouchabac, S., et al. (2020). Virtual reality (VR) in assessment and treatment of addictive disorders: A systematic review. *Front. Neurosci.* 13, 1409. doi:10.3389/fnins.2019.01409

Sevigny, J., Chiao, P., Bussière, T., Weinreb, P. H., Williams, L., Maier, M., et al. (2016). The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. *Nature* 537, 50–56. doi:10.1038/nature19323

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., et al. (1998). The mini-international neuropsychiatric interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 59 (2), 22–33. quiz 34-57.

Slater, M., and Bouchard, S. (2004). Presence and emotions. *Cyberpsychol. Behav.* 7, 121–123. doi:10.1089/109493104322820200

Slater, M., and Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Front. Robot. AI* 3, 1–47. doi:10.3389/frobt.2016.00074

Spreij, L. A., Visser-Meily, J. M. A., Sibbel, J., Gosselt, I. K., and Nijboer, T. C. W. (2020). Feasibility and user-experience of virtual reality in neuropsychological assessment following stroke. *Neuropsychol. Rehabil.* 32, 499–519. doi:10.1080/09602011.2020.1831935

US Food and Drug Administration (2020). DDT COA #000107: Virtual reality functional capacity assessment tool (VRFCAT). Available at: https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/ddt-coa-000107-virtual-reality-functional-capacity-assessment-tool-vrfcat (Accessed January 16, 2022).

Valmaggia, L. R., Freeman, D., Green, C., Garety, P., Swapp, D., Antley, A., et al. (2007). Virtual reality and paranoid ideations in people with an "at-risk mental state" for psychosis. *Br. J. Psychiatry* 51, s63–s68. doi:10.1192/bjp.191.51.s63

Van Der Linden, W. J. (2006). A lognormal model for response times on test items. *J. Educ. Behav. Statistics* 31, 181–204. doi:10.3102/10769986031002181

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence (WASI)*. San Antonio, Texas: Psychological Corporation.

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

Wilson, R. S., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., et al. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychol. Aging* 17, 179–193. doi:10.1037/0882-7974.17.2.179

Wolf, T. J., Edwards, D. F., and Giles, G. M. (2019). *Functional cognition and occupational therapy: A practical approach to treating individuals with cognitive loss*. AOTA Press. doi:10.7139/2017.978-1-56900-477-7