Detecting cognitive impairment through an age-friendly serious game: The development and usability of the Spatial Performance Assessment for Cognitive Evaluation (SPACE)

Colombo, Giorgio; Minta, Karolina; Grübel, Jascha; Tai, Wei Lin Eunice; Hölscher, Christoph; Schinazi, Victor R.

Published in:
Computers in Human Behavior

DOI:
10.1016/j.chb.2024.108349

Computers in Human Behavior 160 (2024) 108349

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Detecting cognitive impairment through an age-friendly serious game: The development and usability of the Spatial Performance Assessment for Cognitive Evaluation (SPACE)

Giorgio Colombo[a,b,*,1], Karolina Minta[a,c,1], Jascha Grübel[d,e], Wei Lin Eunice Tai[a], Christoph Hölscher[b], Victor R. Schinazi[a,f]

[a] Future Health Technologies, Singapore-ETH Centre, Campus for Research Excellence and Technological Enterprise (CREATE), Singapore 138602, Singapore
[b] Cognitive Science, ETH Zürich, Zürich, 8092, Switzerland
[c] Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117600, Singapore
[d] Civil, Environmental and Geomatic Engineering, ETH Zürich, Zürich, 8093, Switzerland
[e] Laboratory of Geo-information Science and Remote Sensing, Wageningen University & Research, Wageningen, 6700, the Netherlands
[f] Department of Psychology, Bond University, Robina, 4226, Queensland, Australia

ARTICLE INFO

Handling Editor: Dr Andrew Lepp

Keywords:
Ageing
Cognitive impairment
Digital assessment
Serious games
Spatial ability
Usability

ABSTRACT

The Spatial Performance Assessment for Cognitive Evaluation (SPACE) is a novel iPad serious game designed to identify differences in spatial ability indicative of early signs of cognitive impairment. This paper reports on the development of SPACE and presents the results from three usability studies across different ages. Study 1 compared the traditional tap and swipe control interface with a semiautomated interface. Study 2 investigated the benefits of a UI widget that displayed the rotation performed in the Virtual Environment (VE). Study 3 evaluated the effects of a simplified configuration of landmarks explored in the VE. Findings across the studies indicated that age was the primary factor influencing performance. Younger participants consistently outperformed older ones across various tasks and reported higher usability and lower workload. Despite notable performance improvements for the tasks in SPACE, the new control interface, UI widget, and simplified configuration only had a minimal impact on usability. Younger participants rated SPACE above the level of mature products, while older participants found it useable but not always engaging. Critically, the significant interactions between age and experimental conditions indicated that younger and older participants benefited differently from the design modifications. Here, the semi-automated control, the simplified configuration, and, to a lesser extent, the UI widget showed promise in mitigating age-related performance differences while maintaining the level of challenge necessary to assess differences in cognitive status. This research showcases the potential of SPACE as a serious game and emphasises the importance of balancing simplicity with task demands for future unsupervised deployment.

1. Introduction

According to projections by the World Health Organization (WHO), a staggering 2.1 billion individuals will be 60 years or over by 2050 (WHO; World Health Organization, 2023). As the population continues to age, the increasing prevalence of medical conditions such as dementia will impose a greater burden on healthcare systems around the globe. Dementia is a degenerative neurological syndrome characterised by the deterioration of cognitive functions, including memory, reasoning, language, and spatial ability (Jack & Holtzman, 2013). Currently, dementia impacts around 55 million people worldwide and constitutes one of the major causes of dependency and death among older adults (World Health Organization, 2023). The WHO predicts that by 2050, approximately 139 million individuals will be living with dementia, and the direct global expenditure on dementia care is expected to reach $2.4 trillion (Velandia et al., 2022). Indeed, dementia poses a multifaceted challenge to the healthcare system, including demand for specialised long-term medical care, caregiver support, and significant healthcare

* Corresponding author. Future Health Technologies Singapore-ETH Centre, 1 Create Way CREATE Tower, #06-01, 138602, Singapore.
E-mail address: giorgio.colombo@sec.ethz.ch (G. Colombo).
[1] These authors equally contributed to this paper.

https://doi.org/10.1016/j.chb.2024.108349
Received 16 February 2024; Received in revised form 10 May 2024; Accepted 13 June 2024
Available online 17 June 2024

costs associated with diagnosis, treatment, and the ongoing management of the disease. Dementia also poses severe and often underestimated indirect societal and personal costs related to informal caregiving (Wong, 2020). With the absence of a known cure for dementia, it is critical to implement early and targeted interventions that can help slow down the progression of the disease and support individuals in upholding their autonomy and quality of life. To fully capitalise on the advantages offered by early interventions, it is essential to have a diagnosis that is both timely and accurate.

Dementia is typically preceded by a preclinical stage known as Mild Cognitive Impairment (MCI), in which cognitive decline is present but not sufficiently severe to disrupt daily activities (Morris et al., 2001; Petersen et al., 2001). Indeed, MCI symptoms can already be present when individuals are in their 40s (Singh-Manoux et al., 2012). Since the pathophysiological processes of more advanced forms of dementia like Alzheimer's Disease (AD) initiate years before symptoms manifest, MCI presents a propitious stage for the early detection and the deployment of interventions when they are more likely to succeed. Researchers are now dedicating extensive efforts towards identifying biomarkers for cognitive impairment (Blennow et al., 2015; Chong et al., 2021; Soldan et al., 2020). The analysis of Cerebrospinal Fluid (CSF) and Positron Emission Tomography (PET) are reliable methods for detecting pathophysiological changes related to dementia. However, their invasiveness, high costs, and limited availability often hinder their use in the early diagnosis. Neuropsychological tests can be used for cognitive evaluation at a lower cost but require a trained neuropsychologist and can take hours to administer. While more rapid cognitive screening tests like the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) and the Mini-Mental State Examination (MMSE; Folstein, 1983) serve as alternatives to identifying advanced dementia cases, their sensitivity to detect early signs of cognitive impairment can be limited (Arevalo-Rodriguez et al., 2013; Dautzenberg et al., 2019; McLennan et al., 2011). Notably, existing cognitive assessments primarily focus on memory deficits and tend to overlook other domains affected by MCI, such as spatial abilities.

This lacuna is surprising because MCI and AD adversely affect the hippocampus and entorhinal cortex (Devanand et al., 2007; Pennanen et al., 2004) through extracellular plaque deposits of the β-amyloid (Aβ) peptide (Chételat et al., 2009) and neurofibrillary tangles of the hyperphosphorylated tau protein (Jack et al., 2010). Critically, these two brain regions are also known to play an instrumental role in the spatial coding of locations (O'Keefe & Nadel, 1979) and in assisting individuals to keep track of changes in their position and orientation during navigation (Hafting et al., 2005; Jack et al., 2018; McNaughton et al., 2006; Schinazi et al., 2013; Schinazi & Thrash, 2018). Indeed, patients suffering from MCI and AD experience a significant decline in spatial abilities in addition to other cognitive deficits (Benke et al., 2014; Coughlan et al., 2018; Hort et al., 2007; Lester et al., 2017; Plácido et al., 2022). Here, atrophy and abnormal neural activity in the hippocampus and entorhinal cortex of these patients have been associated with reduced performance in a variety of spatial tasks compared to healthy controls (Li & King, 2019). For example, deIpolyi et al. (2007) compared the performance of MCI and AD patients with healthy controls in a series of spatial tasks after they learned a route in the real world. They found that these patients underperformed in the spatial tasks relative to healthy controls and that their performance was associated with atrophy of the hippocampus. Similarly, Schöberl et al. (2020) injected participants with a radiopharmaceutical marker and asked them to complete a series of spatial tests in a complex and unknown environment before undergoing a PET scan. They found that amyloid-positive MCI patients performed worse than controls in the spatial tasks and showed reduced activation in the hippocampus and parietal cortex. Interestingly, these authors report that tests typically used to screen for dementia (e.g., word-list learning, figural learning, and trail-making) were incapable of discriminating between MCI and control participants.

While an in-depth evaluation of spatial ability may improve the sensitivity and specificity of existing cognitive screening tools, navigation assessments in the real world take longer to administer and are particularly problematic for participants with reduced mobility. Recent advancements in Virtual Reality (VR) have opened new avenues for developing spatial tests that provide lifelike environments and interactions suitable for all ages. VR capitalises on the benefits of gamification by actively engaging participants, making it an effective tool for assessments. In addition, advances in mobile technologies now allow for VR tools to be easily deployed in portable devices, further facilitating testing outside the laboratory or clinic. Researchers have already started to uncover the potential of VR for assessing various cognitive functions (Groppell et al., 2019; Meier et al., 2021), including spatial and navigation abilities compromised by pathological ageing (Chan et al., 2016; Cushman et al., 2008; Diersch & Wolbers, 2019; Plácido et al., 2022; Puthusseryppady et al., 2022; Tu et al., 2015). However, VR assessments for cognitive evaluation must go beyond creating engaging experiences and provide accurate, inclusive, user-friendly, and secure tools. As such, it is vital that researchers prioritise usability when developing digital technologies, especially when designing for use by older adults.

In this paper, we present the Spatial Performance Assessment for Cognitive Evaluation (SPACE). SPACE is a novel serious game designed to identify deficits in spatial ability indicative of early signs of cognitive impairment. In SPACE, players assume the role of an astronaut on a mission to explore an unfamiliar planet by completing a series of spatial tasks. Each task is specifically designed to probe different aspects of their spatial abilities. In three studies, we evaluated the usability of different features of SPACE in individuals aged between 21 and 76 years of age in terms of task performance and self-report ratings. In Study 1, we assessed the advantages of a new control interface for navigating the Virtual Environment (VE) designed to minimise interaction with the tablet. In Study 2, we investigated the benefits of a navigation aid in the form of a UI widget that displays the amount of rotation performed by the player in the VE. In Study 3, we examined the usability gains of reducing task complexity relative to the number of landmarks in the VE. To anticipate, we found that the three usability changes improved task performance, but this was not always reflected in self-report ratings. Older participants demonstrated a notable preference and advantage for using the new control interface, whereas the navigation aid proved more beneficial for younger adults. In general, the simplified configuration was capable of mitigating age-related performance differences while maintaining the level of challenge necessary to assess differences in cognitive status.

2. Literature review

2.1. Digital technology use among ageing adults

The recent uptake in the adoption of digital technologies among ageing adults provides a promising backdrop for using VR in cognitive screening. In the last decade, the ownership of mobile devices has risen sharply (Faverio, 2022; Gilbert, 2020). Smartphone ownership has surged among individuals aged 18 to 29 (66%–96%), 30 to 49 (59%–95%), 50 to 64 (34%–83%), and those aged 65 and over (13%–61%). Tablet ownership has also risen among those aged 18 to 29 (16%–46%), 30 to 49 (20%–61%), 50 to 64 (13%–53%), and those aged 65 and over (6%–44%). This trend has been accompanied by a surge in health apps, with an estimated 350,000 healthcare apps in the market (Mathews et al., 2019). Health apps have the potential to disrupt clinical care by offering remote access, minimising the need for frequent visits to the clinic while reducing the costs and burden of travel. This improved accessibility can facilitate repeated testing and the sharing of health data based on weekly or daily assessments (Amagai et al., 2022; Czaja et al., 2013). Altogether, health apps may enable individuals to take an active role in managing their health and enhancing their well-being (Czaja, 2019).

Despite the significant advantages of digital technologies, developing user-friendly apps for older adults is challenging due to the diverse and complex issues that often accompany ageing (Adcock et al., 2020; Czaja et al., 2013). For example, cognitive impairment affects problem-solving skills and information processing and can result in lower attention when working on complex tasks (Gamberini et al., 2006). In addition, reductions in motor skills, including slower response times and limited flexibility in movement, can directly impact the ability of older adults to interact with input devices (Czaja, 2019; Gerling et al., 2012; Rogers & Fisk, 2010). Here, the incidence of chronic conditions (e.g., arthritis) and visual loss further affect the ability of these individuals to manipulate, read and interpret visual information on screens (Kappen et al., 2019). These challenges can generate technology anxiety in an ageing population (Rogers & Fisk, 2010), leading to frustration and higher disengagement rates. Indeed, a 2022 report by Statista revealed that health apps continue to struggle to maintain user engagement, with retention rates as low as 3.7% across all age groups (Ceci, 2023). A review by Amagai et al. (2022) further corroborates these findings, noting the lack of support features, technical difficulties, and overall usefulness of the app as critical factors related to retention. Notably, researchers have found that middle-aged and older adults are more willing to engage with the technology and perceive it as useful when they find it easy to understand its value, interact with the controls (Wiemeyer & Kliem, 2011), and adapt to it (Hamid et al., 2023; Heart & Kalderon, 2013; Peek et al., 2014). In general, younger and middle-aged adults have lower anxiety compared to older adults (Czaja et al., 2006), but in-app training and the inclusion of game elements have been found to significantly enhance user experience (Czaja et al., 2013; Koivisto & Hamari, 2019; Lumsden et al., 2016).

2.2. Gamification and serious games

Gamification and serious games aim to enhance user motivation, increase activity, promote social interaction, and improve quality and productivity (Hamari et al., 2014; Riva et al., 2012). Gamification uses game design elements (e.g., rewards, storytelling) in non-game contexts. In contrast, serious games represent fully developed gaming experiences intentionally designed for non-entertainment purposes (Deterding et al., 2011). Unlike gamification, which selectively incorporates game-related components, serious games immerse users in a complete virtual gaming environment, potentially providing a deeper level of engagement and immersion over extended periods (Ryan et al., 2006). Researchers are actively exploring gamification and serious games in healthcare (Konstantinidis et al., 2017; Pereira et al., 2014; Sardi et al., 2017) to promote physical and cognitive activity (Bamidis et al., 2015; Gamberini et al., 2006; Pedroli et al., 2018), behavioural changes (Alahäivälä & Oinas-Kukkonen, 2016; Baranowski et al., 2008), and engagement and adherence to treatment (Amagai et al., 2022; Dias et al., 2018).

Recent studies have shown that approximately 65% of Americans play video games (Entertainment Software Association, 2023). Interestingly, 45% of Americans aged 50 and above report playing games at least once a month (Kakulla, 2023), and similar figures have also been reported for European countries (Clement, 2023). Playing games offers numerous benefits, including increasing self-confidence, diminishing feelings of loneliness, and stimulating stronger connections with family members (Granic et al., 2014). Games can also enhance cognitive control (Anguera et al., 2013), executive functions (Basak et al., 2008; McCord et al., 2020), processing speed (Nouchi et al., 2012), hand-eye coordination (Rutkowski et al., 2021), visual processing (Belchior et al., 2019), and attention (Abbott, 2013; Latham et al., 2013; Spence & Feng, 2010). Finally, gamified apps and serious games can add a layer of personalisation and engagement to health apps by offering solutions custom-tailored to individual needs, resulting in potentially more targeted and impactful interventions (Gerling & Masuch, 2011). Unfortunately, many commercially available serious games are not well-suited for middle-aged and older individuals, and the vast majority of studies have focused on evaluating the effects of gamification among young adults (Marston, 2013), neglecting the needs of an ageing population (Havukainen et al., 2020; Martinho et al., 2020).

Several authors have highlighted the intricacies of designing games for an ageing population (Cota & Ishitani, 2015; Czaja et al., 2013, 2019; Gamberini et al., 2006; Marston, 2013; Martinho et al., 2020; Seçer & Us, 2023). Games for this demographic need to be visually appealing, adaptable to various motor skills, and present an appropriate level of challenge (Gerling et al., 2012). Indeed, a recent study targeting middle-aged adults highlighted several key features that make serious games appealing to this age group (Machado et al., 2018). These features include frequent feedback, opportunities for progression and incremental learning, a fun and relaxing gameplay experience, and the elimination of distractions. Similarly, to assist older adults in feeling more comfortable with games, developers should introduce simplified control interactions (Cota & Ishitani, 2015; Wiemeyer & Kliem, 2011) with in-app training (Czaja et al., 2019), intuitive user interfaces (e.g., zooming function; Czaja et al., 2013; Ijsselsteijn et al., 2007), meaningful goals (Flores et al., 2008) that are accompanied with clear instructions (Czaja et al., 2019; e.g., videos, popup menus; Havukainen et al., 2020) and encouraging feedback (e.g., visual progress indicators, game rewards; Barnard et al., 2013; Flores et al., 2008; Lim et al., 2012). Gamification for middle-aged and older adults should also leverage the potential of storytelling to further engage individuals during repetitive tasks that are essential for accurate health evaluations (Czaja et al., 2019).

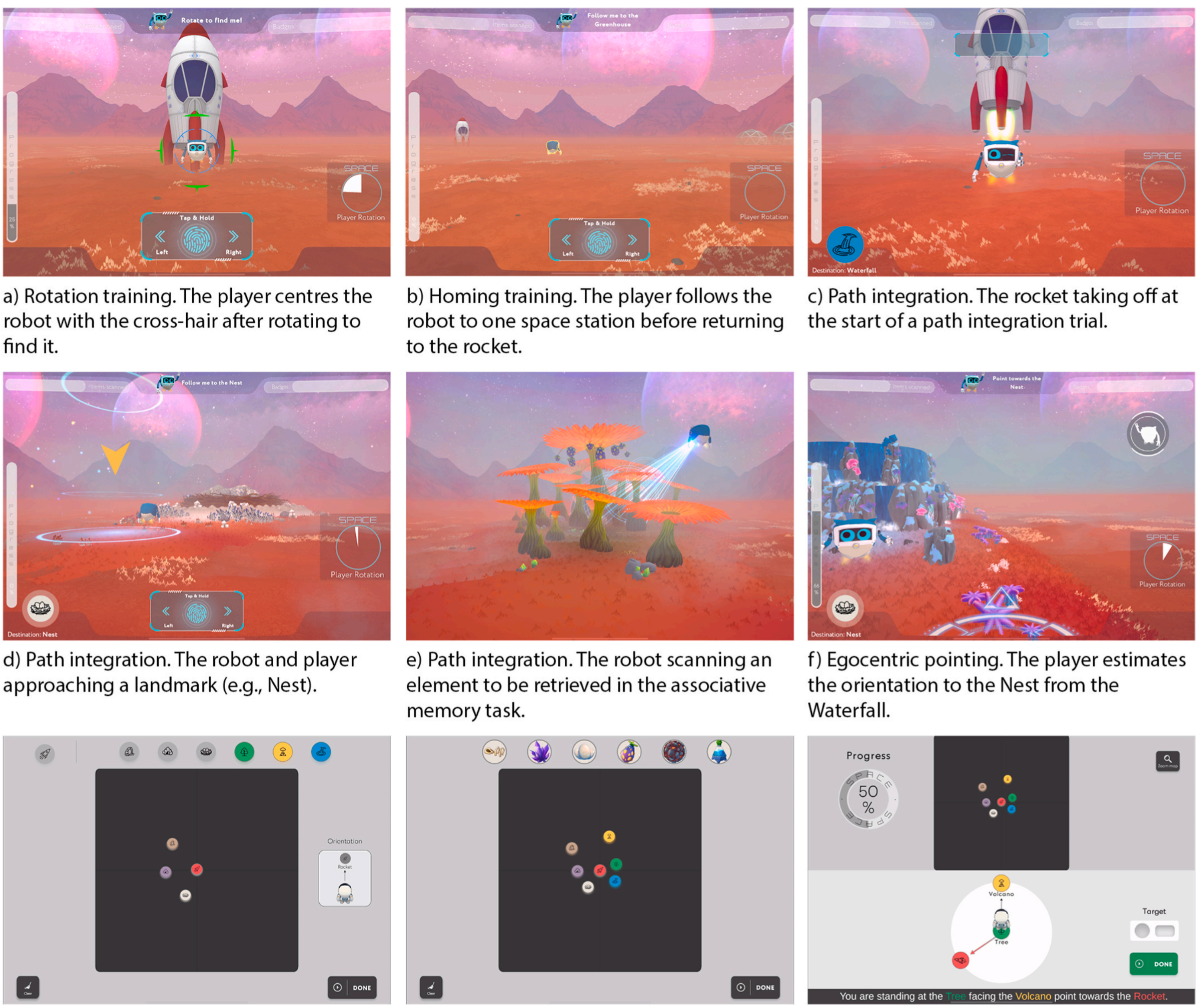
2.3. Serious game and cognitive status

Gamified apps and serious games can play a pivotal role in promoting the health and well-being of older adults at risk of or already experiencing cognitive impairment (Wilson et al., 2022). Several studies have investigated the usability and effectiveness of gamified assessment of cognitive status (Ben-Sadoun et al., 2016; Manera et al., 2015; Tziraki et al., 2017) with varying results. In a study by Manera et al. (2015), patients with cognitive impairment reported positive usability feedback (e.g., motivation, fatigue, and satisfaction levels) when playing a serious game developed to assess and stimulate executive functions. Similarly, other authors (Ben-Sadoun et al., 2016; Tziraki et al., 2017) found positive usability scores and good levels of enjoyment for cognitively intact and impaired participants when playing a cognitive training game. These positive results are further supported by the low number of dropouts during cognitive training studies (Contreras-Somoza et al., 2021). Despite the positive effects highlighted above, gamified cognitive assessment and training technologies have not always been successful or well-received (Bahar-Fuchs et al., 2013). Here, Robert et al. (2014) emphasise the need for more significant efforts in usability testing to create user-friendly interfaces for serious games. Specifically, these authors urge researchers and developers to prioritise ease of use for players and the individuals administering the assessment (e.g., clinicians) who may not have experience with the technology. In addition, the study underscores the lack of standardised methodologies in serious game research and advocates for feasibility and efficacy studies involving users and professionals prior to deployment.

Currently, there exists a multitude of apps purposefully built for the assessment and training of cognitive abilities (Bang et al., 2023; Berg et al., 2018; Berron et al., 2024; Groppell et al., 2019; Lumsden et al., 2016; Maggio et al., 2023; Meier et al., 2021; Vyshedskiy et al., 2022; Öhman et al., 2021). For example, the MoCA DUO (https://mocacognition.com/digitaltools/) is a digital version of the MoCA that offers automated scoring, multilingual support, and interactive elements like on-screen drawing and instructions, simplifying deployment in clinics or via video conferencing. While the MoCA DUO has been effective in replicating the original paper and pencil assessment (Berg et al., 2018; but see, Wallace et al., 2019), it still lacks gamification elements, requires administration by a trained professional, and largely overlooks other cognitive domains like spatial ability. BrainCheck is another digital assessment that has shown promise in discriminating between healthy, MCI, and AD individuals (Groppell et al., 2019). However, similar to the MoCA, BrainCheck relies on typical psychometric tests (e.g., Stroop, trail making, delayed recognition) without considering deficits in spatial ability. Two recently developed gamified assessments, Altoida (Meier et al., 2021) and BrainTrack (https://www.dementia.org.au/braintrack-app), now offer a suite of VR and Augmented Reality (AR) tasks, some of which specifically target deficits in spatial ability. However, the interaction with the AR spatial tasks in Altoida is still somewhat cumbersome, and, to our knowledge, there are no studies on the usability and efficacy of BrainTrack.

Over the last two decades, researchers have attempted to use different aspects of gamification for the design of spatial and navigation tasks in VR aimed at detecting cognitive deficits. These tasks have predominantly focused on one or multiple facets of navigation, including maze learning (Migo et al., 2016; Morganti et al., 2013; Weniger et al., 2011; Zhang et al., 2021), route learning (Levine et al., 2020; Morganti et al., 2013; van der Ham et al., 2020; Weniger et al., 2011; Wiener et al., 2020; Zakzanis et al., 2009), landmark recognition (van der Ham et al., 2020; Zakzanis et al., 2009), path integration (Howett et al., 2019; Tu et al., 2015), map building (Levine et al., 2020; van der Ham et al., 2020), perspective taking (Chan et al., 2016; van der Ham et al., 2020; Wiener et al., 2020), and egocentric/allocentric coding (Castillo Escamilla et al., 2023). For the most part, these tasks have been capable of discriminating between participants who are healthy, cognitively impaired and suffering from advanced types of dementia, albeit with varying levels of success. More recently, Zygouris et al. (2022) used a supermarket test to probe visual and verbal memory, executive function, attention, and spatial navigation ability by asking participants to retrieve items from a shopping list and pay for them by calculating the correct amount. Although the study did not include healthy controls without Subjective Cognitive Decline (SCD), the authors found that participants with SCD and MCI rated the game as highly useable (SUS = 83.11) and that usability scores were correlated with task performance.

To date, Sea Hero Quest (Coutrot et al., 2018) is the most impressive attempt to assess cognitive functioning through a serious game fully



a) Rotation training. The player centres the robot with the cross-hair after rotating to find it.

b) Homing training. The player follows the robot to one space station before returning to the rocket.

c) Path integration. The rocket taking off at the start of a path integration trial.

d) Path integration. The robot and player approaching a landmark (e.g., Nest).

e) Path integration. The robot scanning an element to be retrieved in the associative memory task.

f) Egocentric pointing. The player estimates the orientation to the Nest from the Waterfall.

g) Mapping. The player drags and drops icons in the canvas to create a map of the environment.

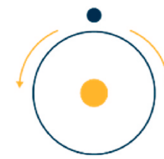
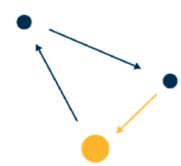
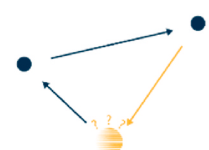
h) Associative memory. The player drags and drops the icons of the elements on the corresponding icon of the landmarks.

i) Perspective taking. A trial of the perspective taking task before confirming the bearing and receiving feedback.

Fig. 1. Screenshots of the training phase and the five spatial tasks in SPACE.

focused on spatial navigation. Despite the game's widespread distribution, its effectiveness in identifying early signs of cognitive impairment has been limited to associations between individual performance and the genetic risk of developing the disease (e.g., APOE-ε4; Coughlan et al., 2019). One major drawback with existing VR spatial assessments is the lack of consensus regarding task development and deployment. Critically, past research lacks standardisation and often uses different tests and different versions of the same tests to probe deficits in navigation. This methodological discrepancy hinders comparison across various spatial tasks and compromises the longevity of the assessments. To effectively assess cognitive status, a comprehensive, engaging, and standardised suite of tests suitable for use in homes, laboratories, or clinical settings is essential. To address this gap, we have developed SPACE as a serious game with a battery of spatial tasks that can be flexibly deployed to assess cognitive status.

Table 1
Overview of training and tasks in SPACE.

Training	**Rotation**	The player learns to rotate in the VE by swiping left and right on the screen to find the robot until it aligns with the centre of a cross-hair.	
	Translation	The player learns to move forward to reach the robot by tapping and holding a button on the screen.	
	Circuit	The player learns to integrate rotations and translations by following the robot in a circuit around the planet.	
	Homing	The player follows the robot from the rocket to two space stations on the planet before being asked to return unaided to the rocket. Throughout this task, the rocket remains visible to the player. Homing introduces the player to the logic of the path integration task.	
Assessment	**Path integration**	The player follows the robot to two distinct landmarks on the planet before being asked to return unaided to the rocket. At each landmark, the robot scans an item that the player will be asked to recall in a subsequent task of the game. Different from Homing in training, the rocket takes off at the start of each trial and remains invisible until the player completes the return journey and signals for its landing. At the end of each trial, the player is transported to the correct location of the rocket.	
	Egocentric pointing	The player performs a memory test for the locations encountered during the path integration task. The player is positioned in front of a landmark or the rocket and asked to complete a series of pointing trials to different landmarks. For each set of trials, the player is teleported to a new landmark on the planet.	
	Mapping	The player is asked to create a map of the planet by dragging and dropping multiple icons representing the landmarks they encountered during the path integration task. After dragging all the icons, an animation showing the correct position of the landmarks is displayed.	
	Associative memory	The player is asked to drag and drop icons in order to pair the items scanned by the robot with the corresponding landmarks they encountered during the path integration task. After completing the task, an animation shows the correct pairings.	
	Perspective taking	The player is asked to imagine standing at a landmark facing another landmark by looking at the map of the planet. The player is then required to indicate the correct bearing towards a third landmark.	

3. SPACE

3.1. The game

SPACE is an iPad-based serious game that provides an in-depth evaluation of spatial ability that may serve as a promising marker for the early detection of cognitive impairment (Fig. 1). We created SPACE with the dual goal of producing an engaging gaming experience and a rigorous assessment of spatial and navigation abilities. The game is set on a foreign planet inspired by the popular game *No Man's Sky* (Hello Games, 2016). The VE is designed to captivate players while creating a suitable environment void of landmarks, which is essential when assessing the acquisition of spatial knowledge. In SPACE, players assume the role of an astronaut sent on a mission to uncover the potential of the new planet to harbour life. The game is optimised for older adults using tasks that are intuitive and complemented by tutorial videos and real-time instructions. Depending on the task in SPACE, players switch between a first-person perspective, simulating the astronaut's viewpoint from the helmet through a Head-Up Display (HUD), and a 2D representation of a panel within the rocket's cockpit (Fagerholt & Lorentzon, 2009). Throughout the game, the HUD provides players with vital information to accomplish the tasks (e.g., instructions, destinations, items scanned, and progress). A companion robot, known as L15A, guides players through each stage of the game and provides instructions, feedback (e.g., facial expressions), and rewards (e.g., performance badges). We also developed various types of controls (i.e., Tap & Swipe, Joystick, and Anchor) that cater to different experience levels and may assist in overcoming difficulties when interacting with tablets for older adults.

The game starts with an extensive training phase (Video 1) that allows players of all ages to familiarise themselves with the control interface while assessing basic visuospatial skills. The training phase effectively reduces confounds by ensuring all players thoroughly understand the control interface before progressing to the main tasks (Grübel et al., 2017). SPACE includes five spatial tasks (i.e., path integration, egocentric pointing, mapping, associative memory, and perspective taking) designed to recruit critical brain regions involved in spatial navigation (Video 2: Hafting et al., 2005; McNaughton et al., 2006; O'Keefe & Nadel, 1979). Table 1 describes the training phase and the different tasks in SPACE.

3.2. Architecture and ecosystem

The architecture of serious games requires a different approach to classical video games because the focus is on collecting detailed information about the player necessary for accurate assessments. SPACE follows the Design, Experiment, Analyse and Reproduce (DEAR) principle as a guiding framework (Grübel, 2023) to effectively ensure the integrity of the assessment and the reproducibility of results (Fig. 2a). The DEAR principle consists of four components that operate in a cyclical pattern. The *Design* phase provides users with a friendly interface to create different configurations of well-established tests. This step streamlines the setup process and allows users to make necessary adjustments before testing. In SPACE, this is achieved via the Experimenter Menu and the configurational panels for each task. The *Experiment* phase uses an implementation of the "Experiments as Code" to provision, deploy, and execute the experiment while enabling data collection (Grübel, 2022). Here, the training phase and the five spatial tasks constitute the experiment. During gameplay, all the data (e.g., position and orientation, time, coordinates, screen interaction) are collected and stored as a JSON file. The *Analyse* phase provides researchers with the necessary support to manage and monitor ongoing experiments and conduct statistical analyses. This phase offers a comprehensive view of the experiment process and integrates tools for real-time or post hoc analyses. We use a custom-made R script to analyse JSON files and extract individual and group scores from the various tasks in SPACE. The *Reproduce* phase gathers the experiment software, the collected data, and the analysis scripts into long-term repositories. These repositories are then indexed in the experiment database for a subsequent *Design* phase. SPACE is implemented in Unity3D and uses JSON files to exchange information on the experimental design and collected data with trusted remote servers. During game production (Fig. 2b), a GitLab LFS repository supports automatic builds on a Jenkins server and deployment to iOS via TestFlight. Expansions and maintenance can be performed in Jenkins' Continuous Integration pipeline.

The SPACE ecosystem (Video 3) is designed to allow researchers and

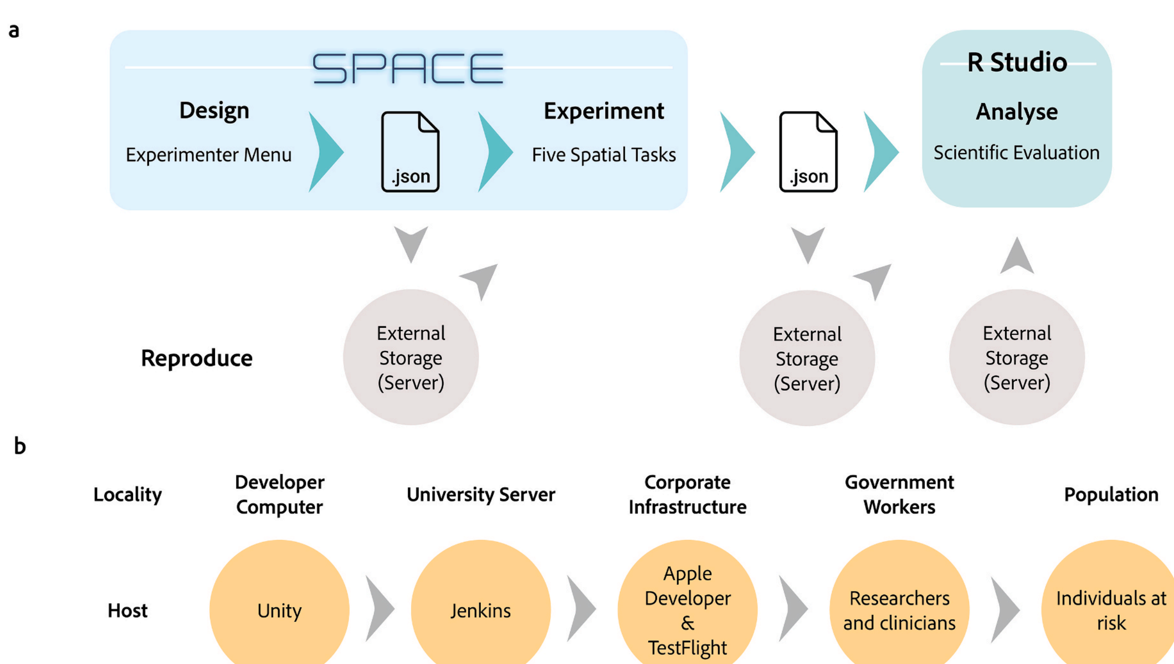
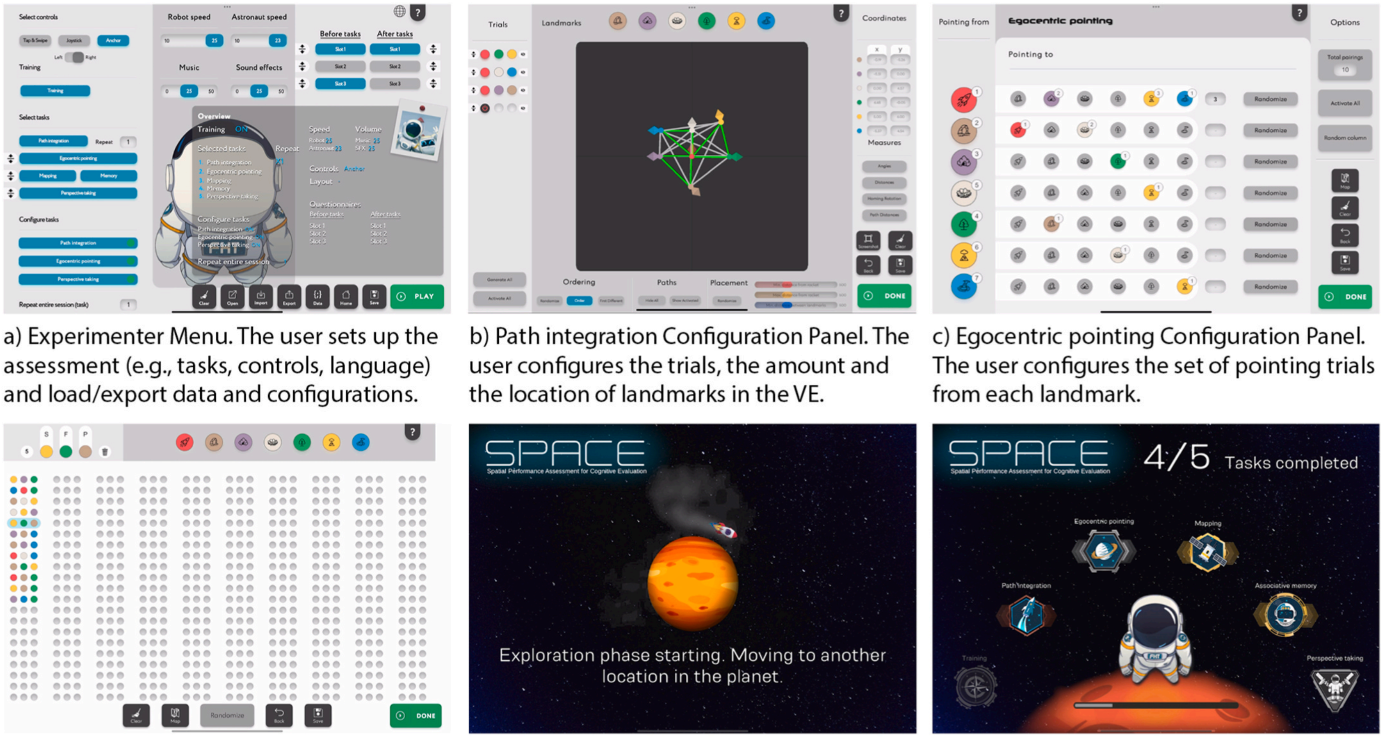


Fig. 2. Architecture and development of SPACE. a) The implementation of the DEAR principle in SPACE. b) The production, distribution, and application pipeline of SPACE.



a) Experimenter Menu. The user sets up the assessment (e.g., tasks, controls, language) and load/export data and configurations.

b) Path integration Configuration Panel. The user configures the trials, the amount and the location of landmarks in the VE.

c) Egocentric pointing Configuration Panel. The user configures the set of pointing trials from each landmark.

d) Perspective taking Configuration Panel. The user defines the trials for the perspective taking task.

e) Transition screen. This screen indicates to the player that training is completed and that the assessment will start.

f) Task loading screen. The player is given feedback on the performance using badges before moving on to the next task.

Fig. 3. Screenshots of the Experimenter Menu, configuration panels, and transition screens in SPACE.

clinicians with limited coding ability to easily configure each task according to their needs (Fig. 3). Through the Experimenter Menu, users can select the language, type of control interface, the speed of the robot and astronaut, set the volume of music and sound effects, and decide whether to include pre-and post-game questionnaires. Users can also decide which spatial tasks (including training) will be part of the assessment, along with the task order and number of session repetitions. Due to the game's inherent logic, the path integration task is always the first to be completed since the knowledge acquired during this task (e.g., the position of landmarks) is necessary to complete the subsequent tasks. Through various configuration menus, users can set the number and position of landmarks in the VE included in the configuration, along with the order and characteristics of each trial. Inherent to the configuration menus, a logic system allows users to proceed only if all the generated trials can be executed in the VE. Other functions such as repetitions, automated trial generation, and randomisation of trials offer novel and quick ways to design and conduct experiments on the fly. Finally, users can verify, in real-time, measures of the angular and distance relationships between landmarks to define different levels of difficulty. SPACE was designed to facilitate deployment in clinical settings where time is of the essence, and the clinician's knowledge of the SPACE ecosystem is limited. Through shortcut buttons or by importing JSON files in the Experimenter menu, clinicians and researchers can quickly load, administer, and share predefined configurations for their assessments consistently across clinics and laboratories worldwide.

4. Materials and methods

4.1. Participants

Two hundred sixty-one healthy participants aged between 21 and 76 years were recruited from the general population through social media platforms (e.g., Facebook, LinkedIn, Telegram) and flyers distributed in the community. Participants with neurological disorders, severe visual impairment or blindness, deafness, a history of seizures, epilepsy, or acute cardiac events were excluded from the study. Due to unforeseen technical issues (e.g., app crashing, refusal to answer a questionnaire), five participants were entirely excluded from the data analysis. In addition, some study participants had incomplete data entries, resulting in missing values in our dataset. Only one participant abandoned testing during the path integration task. These data were omitted from specific analyses, but the participants were not discarded. In total, data from 256 participants were included in the final analyses. In three studies, participants were assigned to four experimental conditions in which we manipulated the control interface, the presence of a UI element, and the complexity of the spatial configuration of the VE used during testing (Fig. 4). The four conditions were carefully balanced in terms of age, ensuring that the proportions of participants from each age group were consistent across all conditions. Table 2 presents the sample characteristics for each condition. Ethical approval to conduct this study was provided by the Parkway Independent Ethics Committee (PIEC/2022/010) and the ETH Zurich Ethics Commission (EK 2021-N-193). Written informed consent was acquired from all participants prior to conducting the study. All studies were performed in accordance with the

Fig. 4. A schematic representation of the conditions tested in each study. Study 1 (S1) tested two different types of control interfaces: Tap & Swipe (T) and Anchor (A). Study 2 (S2) tested the effect of adding a UI element (W) compared to the Anchor condition from Study 1. Study 3 explored the effects of a simplified configuration of landmarks (S) compared to the default configuration with the Widget condition from Study 2.

Table 2
Demographic characteristics of the sample for all studies.

Studies		Study 1: Control Interface			
			Study 2: UI Widget		
				Study 3: Configuration	
	Condition	Tap & Swipe	Anchor	Widget	Simplified
Age groups	n	92	92	36	36
21–29	n	21	20	12	12
30–39		20	20	10	10
40–49		9	10	4	4
50–59		14	13	2	2
60–69		27	25	4	7
70–79		1	4	4	1
Age	Median, Mean ± SD	45.50, 46 ± 16	43, 46 ± 17	34, 40 ± 17	35.50, 40 ± 16
MoCA		27, 27 ± 2	27, 27 ± 2	27, 27 ± 3	27, 27 ± 2
Depression		2, 2.34 ± 1.68	1, 2.40 ± 1.93	2, 2.83 ± 1.76	2, 2.72 ± 1.86
Anxiety		3, 3.17 ± 2.10	3, 3.25 ± 2.11	3, 3.64 ± 2.23	3, 3.47 ± 1.95
Stress		3, 4.17 ± 2.31	3, 3.86 ± 2.21	3, 4.08 ± 2.10	3, 4.42 ± 2.30
Alcohol consumption		1, 2.17 ± 3.22	1, 2.39 ± 5.35	1, 1.28 ± 0.81	1.25, 1.71 ± 1.20
Sleep		6.50, 2.49 ± 0.93	7, 2.49 ± 0.91	7, 2.49 ± 0.84	7, 2.49 ± 0.92
Walking		7, 10.10 ± 10.92	7, 8.92 ± 7.72	7, 9.15 ± 8.71	8, 9.92 ± 9.70
Physical activity		2, 2.68 ± 2.46	2, 3.03 ± 3.49	2.75, 2.96 ± 2.14	2, 2.06 ± 1.70
Tablet expertise		3, 2.49 ± 0.69	3, 2.52 ± 0.67	3, 2.61 ± 0.69	3, 2.44 ± 0.74
Gender	% Male/Female	37%/63%	49%/51%	36%/64%	56%/44%
Education	% High school/University	20%/80%	32%/68%	26%/74%	20%/80%
Handedness	% Right/Left	91%/9%	92%/8%	97%/3%	89%/11%
Vision defects	% Yes/No	67%/33%	72%/28%	67%/33%	58%/42%
Chronic Health conditions	% Yes/No	22%/78%	23%/77%	22%/78%	6%/94%
Navigation training	% Yes/No	9%/91%	9%/91%	3%/97%	8%/92%
Falls	% Yes/No	19%/81%	13%/87%	8%/92%	14%/86%

* Data on traumatic brain injury and smoking were excluded from the analysis since only one participant reported having a traumatic brain injury and only two participants reported being smokers across all conditions.

Declaration of Helsinki.

4.2. Materials

4.2.1. Questionnaires

Participants were asked to complete a series of questionnaires before and after the tasks in SPACE. At the start of the experiment, participants completed a vision test, the MoCA, a sociodemographic and health questionnaire and the Santa Barbara Sense of Direction Scale (SBSOD). Following SPACE, participants completed the System Usability Scale (SUS), the User Experience Questionnaire (UEQ), the NASA Task Load Index (NASA-TLX), the Presence questionnaire, and a Debriefing questionnaire.

Sociodemographic and health questionnaire. The sociodemographic and health questionnaire was used to collect data on the participants' background (e.g., age, gender, education level, handedness, experience with tablet, previous navigation training), health status (e.g., vision defects, chronic health conditions, presence of traumatic brain injury), and psychosocial well-being (e.g., levels of depression, anxiety, and stress in the last six months) and health habits (e.g., smoking, alcohol, falls in the last 12 months, hours of sleep per day, hours of walking and vigorous physical activity per week). In the questionnaire, participants were also required to specify whether they were wearing glasses or contact lenses during the test.

MoCA. The MoCA is a one-page, 30-point test administered in person by a qualified experimenter as a screening tool for cognitive impairment. The MoCA evaluates six cognitive domains: memory, executive function, visuospatial, language, attention, and orientation. A score below 26 points is typically used as a cut-off for MCI. Previous research has shown that the MoCA has a sensitivity of 90% and a specificity of 87% (Nasreddine et al., 2005). A meta-analysis by Pinto et al. (2019) showed that the MoCA surpasses the Mini-Mental State Examination (MMSE) for detecting MCI, with AUCs ranging from 0.71 to 0.99 for the MoCA compared to 0.43 to 0.94 for the MMSE. The MoCA also outperforms the MMSE for the detection of Alzheimer's Disease (AD), with AUCs ranging from 0.87 to 0.99 for the MoCA and 0.67 to 0.99 for the MMSE.

SBSOD. The SBSOD is a 15-item scale developed by Hegarty et al. (2002) to assess self-perceived navigation ability. For each item, participants are asked to report their level of agreement with a series of statements (e.g., *I very easily get lost in a new city*) on a 7-point Likert Scale. A final score is computed by averaging all scores, with greater scores indicating a higher perceived sense of direction. The SBSOD has demonstrated good internal reliability ($\alpha = 0.88$) and high test-retest reliability ($r = 0.91$).

SUS. The SUS is a 10-item self-report questionnaire used for assessing the perceived usability of software, websites, and other interactive systems (Brooke, 1996). For each item in the SUS (e.g., *I found the system unnecessarily complex*), participants are asked to respond using a 5-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree". *SUS scores* range from 0 to 100, with higher scores indicating better usability. Over the years, the SUS has demonstrated good reliability, ranging from $\alpha = 0.85$ to 0.90, and is sensitive to differences among types of interfaces and changes made to a product (Assila et al., 2016).

UEQ. The UEQ is a 26-item self-report questionnaire for assessing the user experience of a product or system (https://www.ueq-online.org/). The UEQ items are grouped into six scales: attractiveness, efficiency, perspicuity, dependability, stimulation, and novelty. Some authors (Laugwitz et al., 2008) suggest considering attractiveness as a measure of pure valence (i.e., related to the overall impression of SPACE) and combining perspicuity, efficiency, and dependability into pragmatic aspects (i.e., related to the quality of the interaction when using SPACE), while novelty and stimulation into hedonic aspects (i.e., related to the perceived pleasure or fun while engaging with SPACE). The UEQ is assessed using a 7-point Likert scale between pairs of polar adjectives (e.g., *Annoying* vs. *Enjoyable*, *Impractical* vs. *Practical*). Scores for the UEQ range from −3 to 3, with higher scores indicating higher usability. The UEQ has demonstrated varying degrees of reliability across different languages and contexts ($\alpha = 0.55$ to 0.95), with lower scores typically associated with the dependability scale (Schankin et al., 2022).

NASA-TLX. The NASA-TLX is a 6-item tool designed for assessing and quantifying subjective workload on a 20-point scale (Hart & Staveland, 1988). The NASA-TLX assesses the mental, physical, and temporal demands, as well as performance, effort, and frustration experienced during a task (e.g., *How mentally demanding was the task?*). The NASA-TLX has a maximum score of 100, with higher scores indicating greater perceived workload. In this study, we administered the unweighted version of the scale without pairwise comparisons. As such, raw ratings from each subscale were averaged into an overall workload score (Hart, 2006). The NASA-TLX has shown good reliability (α > 0.80) for individual items and moderate reliability (α > 0.60) for the overall score (Xiao et al., 2005).

Presence questionnaire. The Presence questionnaire (Witmer & Singer, 1998) consists of 19 questions on a 7-point Likert scale designed to measure the sense of presence experienced in a VE (e.g., *How involved were you in the virtual environment experience?*). The Presence questionnaire evaluates five domains: realism, possibility to act, quality of the interface, ability to investigate, and self-evaluation of performance. We employed the revised version of the questionnaire provided by the Université du Québec en Outaouais Cyberpsychology Lab, which has shown good reliability (α = 0.84). This version has a maximum score of 133 points, with higher scores indicative of greater presence (Witmer et al., 2005).

Debriefing questionnaire. The Debriefing questionnaire consists of three open-ended questions designed to probe what aspects of each task participants liked or disliked, as well as their overall impressions of SPACE.

4.2.2. Hardware and software

SPACE was deployed on a 10.2-inch iPad with Wi-Fi and 256 GB memory running iOS version 16.6.1. A description of all the tasks in SPACE is provided in Table 1. The sampling rate for behavioural data is set to 4 samples per second to be as accurate as possible while respecting the trade-off between data availability and game responsiveness. The vision test was administered using the iPad app MDCalc (https://www.mdcalc.com). Data for all the questionnaires were collected with the iPad using the Qualtrics XM online survey platform (www.qualtrics.com). Gait data was collected using WitMotion sensors (WT901BLECL Bluetooth 5.0 Accelerometer, https://www.wit-motion.com).

4.2.3. Procedure

To evaluate the usability of SPACE, we created four different conditions: Tap & Swipe, Anchor, Widget, and Simplified. We compared pairs of conditions in three studies to systematically assess differences in performance in the various spatial tasks in SPACE and self-reported usability (Fig. 5). The preferred condition of each study was carried forward and compared with a new condition in the subsequent study (Fig. 4). In Study 1, we compared the standard control interface (Tap & Swipe) with a new semi-automated control interface (Anchor) designed to reduce the interactions necessary to manoeuvre around the VE. In Study 2, we evaluated the benefits of a UI element that displayed the amount of rotation performed in the VE (Widget) by comparing it to the preferred condition (Anchor) from the previous experiment that did not include the widget. In Study 3, we examined the effects of a simplified configuration of landmarks in the VE (Simplified) by comparing it to the default complex configuration from the previous experiment (Widget). Given that the paper's primary focus is on task performance and usability, some data collected from other questionnaires (e.g., socio-demographic and health questionnaire, SBSOD) and the gait assessment were excluded from analysis and are reserved for subsequent publications addressing dementia risk factors and the relationship between fall risk, spatial ability, and cognitive status.

The three studies followed the same testing protocol. Upon arriving

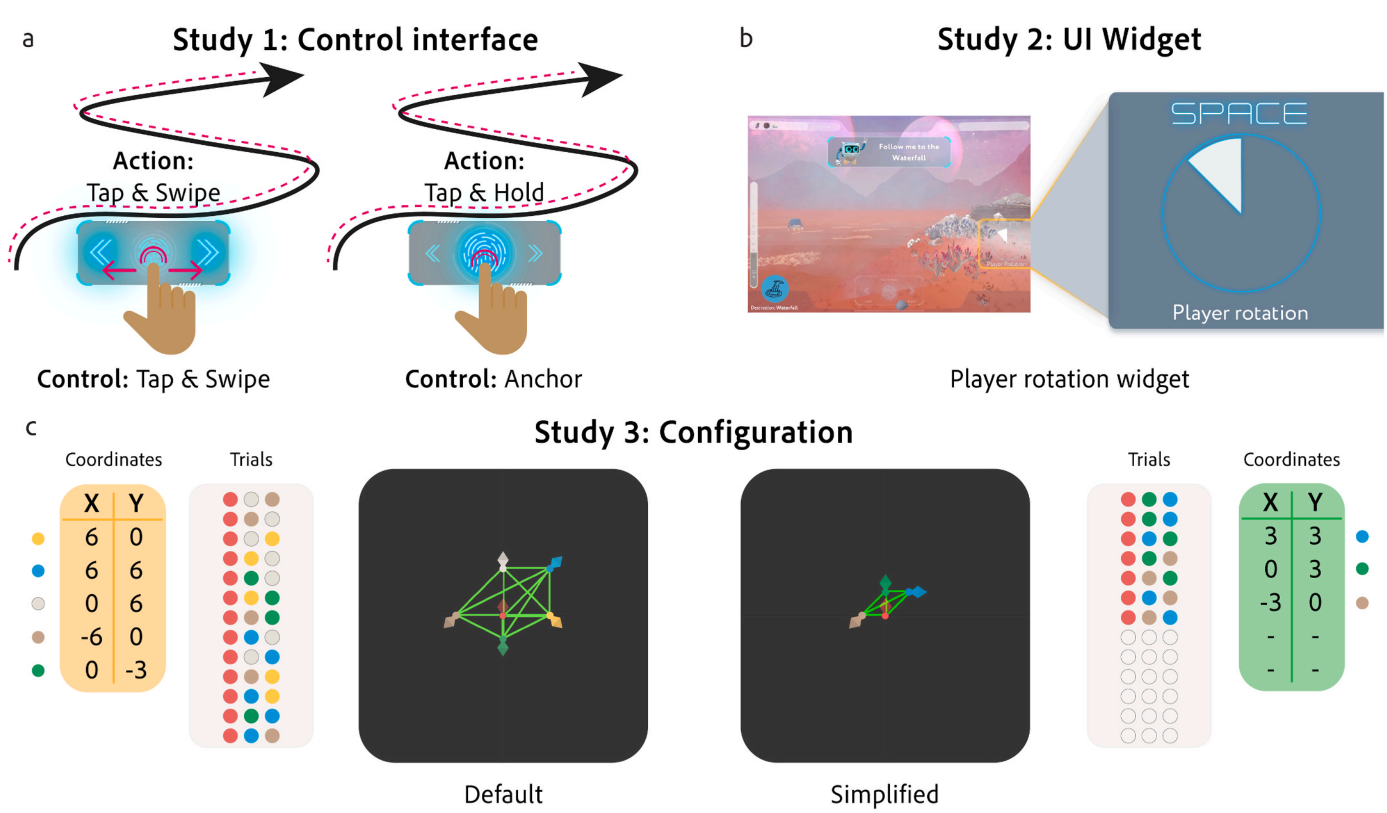


Fig. 5. Schematic representation of the three Studies. a) Study 1: Players used either the Tap & Swipe (left) or the Anchor (right) control to navigate the VE. b) Study 2: Players navigated with or without the support of a widget that indicated the amount of rotation in the VE. The widget during the path integration task (left). A zoomed version of the rotation widget (right). c) Study 3: Players completed SPACE with the Default or a Simplified configuration of landmarks and trials. The tables present the coordinates, list and order of trials.

at the laboratory, participants were introduced to the study procedure and asked to sign the consent form to participate in the study. Once consent was obtained, participants completed the vision test followed by the MoCA. Participants were then asked to complete the sociodemographic and health questionnaire. Depending on the study, participants were randomly assigned to one of two conditions. Instructions for each task were given verbally by a researcher and displayed in a text popup as a game element before and during each task. Following completion of the training phase, participants underwent the five tasks in SPACE. The order of the tasks was not counterbalanced since they followed the logic and story of the game. At the end of the experiment, participants completed the SUS, UEQ, NASA-TLX, the Presence questionnaire, and the Debriefing questionnaire. The final part of the study consisted of a gait assessment with the WitMotion sensors. The questionnaire administration typically spanned 30–40 min, and participants were offered a 5-min break before playing SPACE.

4.3. Design and analysis

4.3.1. Quantitative analysis

For each task in SPACE, we extracted the following performance variables. *Training time* consists of the duration (in seconds) required for each player to complete the Rotation, Translation, and Homing phases of the Training. Here, the time to complete the Circuit stage was excluded because it only applied to the Tap & Hold condition. *Path integration time* (in seconds) refers to the average duration required to finish all path integration trials. *Path integration distance* refers to the average distance between the player's final position and the rocket's original position for all path integration trials. A greater distance indicates a larger error. *Egocentric pointing error* is the average angular deviation (in degrees) between the estimate made by the player and the target landmark for all trials in the task. *Mapping accuracy* was computed using bidimensional regression. Bidimensional regression is a statistical technique for assessing the degree of association (r^2) of two planar configurations (i.e., the real map of the planet and the map created by the player) of related coordinate data (Friedman & Kohler, 2003; Tobler, 1965). The *Associative memory score* was computed as the percentage of correct pairings between scanned elements and landmarks. *Perspective taking error* is the average angular deviation (in degrees) between the estimate made by the player and the target landmark for all trials in the task. For the self-reported usability questionnaires, we computed the SUS scores following the guidelines provided by Brooke (Brooke, 1996). For the UEQ, we extracted separate scores for the *UEQ Attractiveness*, *UEQ Hedonic,* and *UEQ Pragmatic,* following the guidelines provided by Laugwitz et al. (2008). For the NASA-TLX, we used the average score for all scales as a general measure of perceived workload. For the Presence questionnaire, we computed the sum of all individual items, excluding those related to sounds (i.e., items 20–22).

Before conducting inferential statistics, we verified whether our data violated the homoscedasticity and normality assumptions of correlation, Analysis of Variance (ANOVA) and linear regression. When assumptions for homoscedasticity and normality were met, parametric statistics were applied. When the assumptions were violated, we used nonparametric or robust statistics. While having a balanced sample size is not one of the assumptions in linear models, heteroscedasticity becomes more problematic when sample sizes are unbalanced (e.g., Study 2). Here, robust regression mitigates the impact of heteroscedasticity and reduces the undue influence of outliers on the regression estimates by giving them less weight in the model-fitting process (Field & Wilcox, 2017). This method is particularly useful when the smaller samples come from populations with greater variability (older adults).

For each experimental condition, we conducted Spearman's rank-order correlations to assess the convergent and discriminant validity between the tasks in SPACE and between the self-reported usability questionnaires. We also conducted correlations between the SPACE tasks and the self-reported usability questionnaires to assess the concurrent validity between our objective (performance in SPACE) and subjective (self-reported) usability measures. For each of the studies, we conducted between ten (Study 1) and twelve (Studies 2 and 3) separate regression models. The outcome variable for each of these models was the performance score on the different tasks in SPACE or the self-reported usability scores for the SUS, UEQ Attractiveness, UEQ Pragmatic and UEQ Hedonic scales, the NASA-TLX and Presence questionnaire. We entered *Age* and *MoCA* as continuous predictor variables, while *Condition* was entered as a categorical dichotomous predictor variable. Finally, we included an interaction term between *Age* and *Condition* in each model.

In all studies, we excluded the Associative Memory task as an outcome variable from our analysis because of ceiling effects, which limited variability and prevented the models from converging. We also excluded the *Condition* predictor for models with the perspective taking score as the outcome variable. Unlike the other tasks in SPACE, the perspective taking task is a stand-alone task in which participants were provided with a top-down view of the VE and could not directly benefit from changes in the control interface, the widget, and the simplified configuration. The NASA-TLX and Presence questionnaires were administered only starting from Study 2 to further explore other usability aspects of SPACE. Finally, we excluded the model with path integration time in Study 3 as completing the complex configuration was expected to take longer since it required more trials. All statistical analyses were performed using R Studio Version 2023.06.0 + 421 (R Studio PBC, Boston, MA, http://www.rstudio.com). Robust ANOVAs and robust regressions were conducted using the R packages *WRS2* (Mair & Wilcox, 2020) and *robustbase* (Maechler et al., 2023; Todorov & Filzmoser, 2009). The threshold for significance for all tests was set at $p = 0.05$.

4.3.2. Qualitative analysis

Answers to the open questions were analysed using thematic analysis (Braun & Clarke, 2006). This analysis aimed to identify what participants liked or disliked about each of the tasks in SPACE and to gather overall feedback on their experience playing the game. This feedback was gathered for each experimental condition and classified based on three age groups: Young (20–39), Middle-aged (40–59), and Old (60+). Following an inductive approach, the analysis was conducted using the steps proposed by Braun and Clarke (2006). First, the researchers familiarised themselves with the data by repeatedly reading participants' feedback on the three questions. Next, initial codes were applied to the feedback and first-level codes that were similar and shared underlying meaning were grouped into overarching themes. The focus and scope of each theme were then assessed to ascertain their coherence, relevance, and fidelity to the data. Finally, each theme was further refined and given clear and descriptive names. Three researchers independently analysed the data, and a fourth researcher oversaw the independent identification of themes and collaborated on the final refinement and naming of the themes. The final codes and themes emerged from overlapping codes noted by the individual researchers.

5. Results

5.1. Sample characteristics

We conducted separate chi-square, parametric, non-parametric and robust ANOVAs to investigate whether our samples differed across conditions. Our results revealed no significant differences between the four conditions in terms of age (F (3.00) = 2.48, $p = 0.084$), MoCA (F (2.62) = 0.00, $p = 1.000$), depression ($\chi^2(3) = 6.82$, $p = 0.078$), anxiety ($\chi^2(3) = 2.07$, $p = 0.558$), stress ($\chi^2(3) = 2.05$, $p = 0.563$), alcohol consumption (F (2.15) = 0.19, $p = 0.835$), average hours of sleep per day (F (3.00) = 1.99, $p = 0.137$), hours of walking per week (F (3.00) = 0.32, $p = 0.794$), hours of physical activity per week (F (3.00) = 0.49, $p = 0.668$), tablet experience ($\chi^2(3) = 1.61$, $p = 0.656$), gender ($\chi^2(3) =$

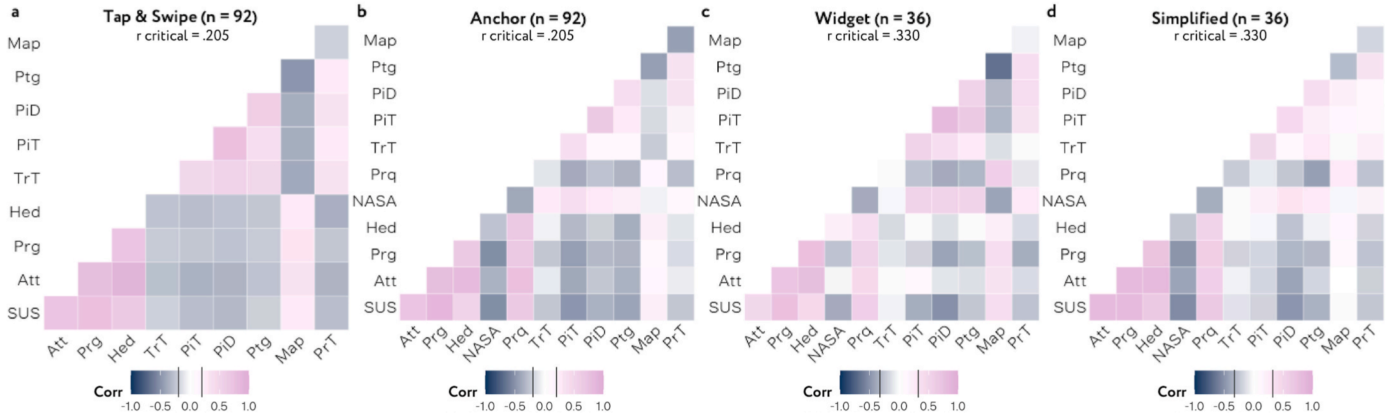


Fig. 6. Correlation matrices for the SPACE tasks and self-reported usability questionnaires for the four experimental conditions. Att = UEQ Attractiveness, Prg = UEQ Pragmatic, Hed = UEQ Hedonic, NASA = NASA-TLX, Prq = Presence questionnaire, TrT = Training time, PiT = Path integration time, PiD = Path integration distance, Ptg = Egocentric pointing, Map = Mapping, PrT = Perspective taking. The vertical black lines in the legend indicate the critical values for r.

5.61, $p = 0.132$), education level ($\chi^2(6) = 9.62$, $p = 0.142$), handedness ($\chi^2(3) = 1.92$, $p = 0.590$), vision defects ($\chi^2(3) = 2.15$, $p = 0.543$), chronic health conditions ($\chi^2(3) = 5.45$, $p = 0.141$), navigation training ($\chi^2(3) = 1.48$, $p = 0.687$), and falls ($\chi^2(3) = 2.55$, $p = 0.466$).

5.2. Convergent, discriminant and concurrent validity

Fig. 6 presents the correlation matrices for the SPACE tasks and self-reported usability questionnaires for the four experimental conditions. Results of the correlations between the five tasks in SPACE revealed low to high convergent validity ($r_{Tap\&Swipe} = 0.20$ to 0.74; $r_{Anchor} = 0.10$ to 0.62; $r_{Widget} = 0.01$ to 0.80; $r_{Simplified} = 0.02$ to 0.46) depending on the experimental condition and the combination of task variables. As expected, we observed higher convergent validity among variables related to the same task (e.g., path integration distance and path integration time) and among variables measuring the same type of spatial ability (e.g., path integration distance and egocentric pointing; egocentric pointing and mapping). Conversely, higher discriminant validity was found between different tasks (e.g., mapping and perspective taking), each designed to assess distinct aspects of navigation ability. Notably, this pattern of validity is more pronounced in the Widget and Simplified conditions.

Results of the correlations between the self-report usability questionnaires also revealed convergent validity ranging from low to high ($r_{Tap\&Swipe} = 0.62$ to 0.94; $r_{Anchor} = 0.26$ to 0.86; $r_{Widget} = 0.04$ to 0.78; $r_{Simplified} = 0.25$ to 0.85) depending on the combination of questionnaire variables. Across all conditions, convergent validity was higher for the variables within the UEQ (e.g., UEQ Hedonic and UEQ Attractiveness), between the SUS and UEQ Pragmatic, between the NASA-TLX and SUS, and between the NASA-TLX and UEQ Pragmatic. Discriminant validity was highest between the NASA-TLX and UEQ Hedonic and Attractiveness subscales. In terms of concurrent validity, the results of the correlations revealed low to medium coefficients ($r_{Tap\&Swipe} = 0.20$ to 0.36; $r_{Anchor} = 0.06$ to 0.48; $r_{Widget} = 0.01$ to 0.56; $r_{Simplified} = 0.00$ to 0.48) between the five tasks in SPACE and the self-report usability questionnaires. Across the four conditions, a discernible pattern emerged in which better task performance was correlated with higher scores in the SUS and UEQ and lower scores in the NASA-TLX. However, differences in the magnitude of these correlations across conditions indicate that this relationship is more stable for the path integration task.

5.3. Study 1: Control interface

In Study 1, we assessed differences in performance and self-reported usability when using two different types of control interfaces (Fig. 5a). The Tap & Swipe control interface offers a default type of interaction in which players are free to move by tapping and swiping on the screen. This interaction can be challenging for some players because it requires them to simultaneously make changes in position and direction. In contrast, the Anchor control is a semi-automated interface in which players can trigger their movement in the VE by simply tapping a button on the screen. Critically, the Anchor control splits the players' response in the path integration task into two steps. Players are required to first estimate the amount of rotation necessary to align with the rocket's position before estimating the distance needed to return to the rocket. This two-step approach simplifies the interaction and may improve performance and perceived usability by providing a more user-friendly way to navigate the VE.

5.3.1. Methods

A power analysis using G*Power (Faul et al., 2007) for a linear multiple regression with four predictors revealed that a sample size of 174 participants was sufficient to detect a moderate effect size ($f^2 = 0.15$), with a power of 0.99 and α set at 0.05. In total, 184 participants were included in this study of which 92 participants were assigned to the Tap & Swipe condition and the other 92 participants to the Anchor condition. The materials and procedure were performed as described in the general Methods section above.

5.3.2. Results

Descriptive statistics for the SPACE tasks and self-report usability questionnaires are presented in Table 3. Results from the regression models are presented in Fig. 7 (task performance), Fig. 8 (self-reported usability questionnaire), and Table S1 in the Supplementary material. Regarding the models with task performance in SPACE as outcome variables, *Age* significantly predicted training time ($\beta = 0.60$, $p < 0.001$), path integration time ($\beta = 0.36$, $p < 0.001$), path integration distance error ($\beta = 0.53$, $p < 0.001$), egocentric pointing error ($\beta = 0.39$, $p < 0.001$), and mapping score ($\beta = -0.35$, $p = 0.013$). For all tasks, performance decreased with age. *Cognitive status* also significantly predicted error scores in path integration distance ($\beta = -0.13$, $p = 0.029$) and perspective taking ($\beta = -0.26$, $p < 0.001$). Here, lower MoCA scores were associated with worse performance. *Condition* was also a significant predictor of training time ($\beta = 0.19$, $p < 0.001$), path integration time ($\beta = -0.17$, $p = 0.005$) and path integration distance error ($\beta = -0.06$, $p = 0.012$). As expected, participants using the Anchor control were faster when completing the training and path integration tasks and committed lower distance errors in path integration compared with those using the Tap & Swipe control. Finally, the interaction between *Age* and *Condition* also significantly predicted training time ($\beta = -0.41$, $p = 0.001$; Fig. 7b), path integration time ($\beta = -0.21$, $p = 0.001$; Fig. 7c), and path integration distance error ($\beta = -0.27$, $p = 0.012$; Fig. 7d).

Table 3
Descriptive statistics for the SPACE tasks and the self-report questionnaire for each condition of the three studies.

Studies		Study 1: Control Interface			
			Study 2: UI Widget		
				Study 3: Configuration	
Condition		Tap	Anchor	Widget	Simplified
n		92	92	36	36
Outcomes	Age	Median, Mean ± SD			
Training time	Young	214.72, 214.82 ± 23.78	236.44, 239.08 ± 27.04	230.71, 233.72 ± 16.37	232.99, 237.32 ± 24.52
	Middle-aged	247.82, 252.42 ± 40.71	249.08, 245.01 ± 34.09	244,50, 255.16 ± 29.16	235.05, 237.91 ± 12.30
	Old	265.32, 261.51 ± 58.87	249.47, 261.29 ± 40.77	264.982, 271.33 ± 41.23	272.98, 277.36 ± 39.44
Path integration time	Young	89.84, 102.45 ± 30.01	95.03, 98.72 ± 15.85	90.73, 94.45 ± 16.40	75.61, 80.57 ± 25.69
	Middle-aged	118.20, 136.34 ± 54.84	104.82, 110.49 ± 24.05	133.24, 129.47 ± 32.75	68.49, 72.22 ± 12.00
	Old	155.53, 173.97 ± 80.88	116.79, 137.84 ± 101.77	122.54, 134.74 ± 45.51	107.22, 116.85 ± 31.83
Path integration distance	Young	213.90, 213.36 ± 60.44	230.10, 247.82 ± 88.70	239.62, 250.92 ± 142.41	143.34, 157.73 ± 70.71
	Middle-aged	283.66, 309.89 ± 120.69	256.87, 284.39 ± 103.75	331.11, 327.91 ± 92.68	167.26, 156.08 ± 61.53
	Old	330.94, 341.63 ± 103.44	299.31, 296.62 ± 84.73	402.41, 408.95 ± 116.66	178.98, 234.28 ± 134.39
Egocentric pointing error	Young	56.53, 53.16 ± 17.21	56.99, 58.41 ± 17.86	51.28, 51.02 ± 21.57	63.22, 58.87 ± 25.25
	Middle-aged	66.80, 65.69 ± 18.36	69.57, 67.31 ± 19.67	78.78, 66.87 ± 26.58	73.32, 76.03 ± 15.05
	Old	67.57, 70.61 ± 11.47	73.45, 73.90 ± 15.45	75.71, 81.55 ± 18.42	80.27, 83.37 ± 13.20
Mapping score	Young	0.48, 0.50 ± 0.34	0.30, 0.42 ± 0.30	0.58, 0.59 ± 0.34	0.59, 0.63 ± 0.29
	Middle-aged	0.20, 0.26 ± 0.22	0.25, 0.33 ± 0.27	0.30, 0.44 ± 0.38	0.67, 0.65 ± 0.28
	Old	0.20, 0.25 ± 0.23	0.28, 0.33 ± 0.22	0.18, 0.27 ± 0.21	0.71, 0.60 ± 0.33
Perspective taking error	Young	12.36, 18.57 ± 14.27	14.57, 20.66 ± 17.38	13.67, 22.01 ± 21.74	18.75, 22.35 ± 15.80
	Middle-aged	22.95, 26.39 ± 15.97	26.16, 32.41 ± 24.10	10.18, 9.38 ± 2.25	14.16, 22.37 ± 19.94
	Old	20.20, 26.01 ± 18.93	39.41, 43.27 ± 23.86	69.83, 55.69 ± 37.38	47.79, 49.84 ± 34.28
SUS	Young	70.00, 69.15 ± 14.10	67.50, 66.06 ± 14.28	66.25, 66.36 ± 16.65	66.25, 62.73 ± 18.26
	Middle-aged	62.50, 61.52 ± 17.48	60.00, 57.93 ± 17.43	57.50, 57.08 ± 20.58	63.75, 65.42 ± 16.69
	Old	48.75, 55.45 ± 18.80	52.50, 52.59 ± 19.08	50.00, 51.88 ± 9.52	56.25, 54.38 ± 11.93
UEQ Attractiveness	Young	1.67, 1.67 ± 0.78	0.75, 0.83 ± 1.06	0.92, 0.77 ± 0.93	1.17, 1.16 ± 1.02
	Middle-aged	1.33, 0.78 ± 1.43	1.17, 0.86 ± 1.40	1.25, 0.89 ± 1.26	1.42, 1.14 ± 1.01
	Old	0.67, 0.77 ± 1.34	0.50, 0.43 ± 1.60	0.92, 0.31 ± 1.45	0.75, 0.65 ± 0.93
UEQ Pragmatic	Young	1.17, 1.10 ± 0.66	1.00, 1.04 ± 0.74	0.96, 0.91 ± 0.69	1.12, 1.01 ± 0.71
	Middle-aged	0.83, 0.81 ± 0.86	0.75, 0.66 ± 0.98	0.75, 0.064 ± 0.95	0.88, 0.72 ± 0.91
	Old	0.50, 0.62 ± 1.22	0.17, 0.25 ± 1.32	0.25, 0.11 ± 0.86	0.33, 0.32 ± 0.90
UEQ Hedonic	Young	1.38, 1.25 ± 0.86	0.75, 0.67 ± 1.13	0.69, 0.66 ± 0.80	0.94, 0.84 ± 0.79
	Middle-aged	0.88, 0.63 ± 1.33	0.62, 0.73 ± 1.20	1.19, 1.04 ± 1.04	1.12, 0.98 ± 1.14
	Old	0.31, 0.72 ± 1.18	0.38, 0.41 ± 1.13	0.25, 0.33 ± 0.74	0.69, 0.73 ± 0.76
NASA-TLX	Young	/	42.08, 41.18 ± 12.21	38.75, 39.75 ± 11.60	42.08, 42.16 ± 15.49
	Middle-aged	/	50.00, 50.08 ± 15.41	51.67, 51.11 ± 10.82	45.42, 45.83 ± 7.71
	Old	/	48.33, 46.22 ± 11.09	51.67, 51.88 ± 10.12	45.00, 44.17 ± 16.19
Presence questionnaire	Young	/	83.00, 85.83 ± 14.02	85.50, 84.77 ± 12.76	90.50, 88.05 ± 14.95
	Middle-aged	/	80.00, 82.65 ± 14.16	84.00, 81.83 ± 20.10	87.00, 84.67 ± 18.13
	Old	/	79.00, 77.52 ± 14.84	78.50, 77.88 ± 13.35	78.00, 79.38 ± 11.66

* Data from NASA-TLX and Presence questionnaire were only collected in Studies 2 and 3.

Critically, older adults were faster and performed better with the Anchor control, while younger adults were faster and performed better with the Tap & Swipe control.

Regarding the models with self-reported usability as the outcome variables, *Age* significantly predicted scores in the SUS ($\beta = -0.40$, $p < 0.001$), UEQ Attractiveness ($\beta = -0.38$, $p < 0.001$), UEQ Pragmatic ($\beta = -0.27$, $p = 0.012$), and UEQ Hedonic ($\beta = -0.30$, $p = 0.004$) scales. For all scales, self-reported usability decreased with age. In addition, *Condition* significantly predicted scores in the UEQ Attractiveness scale ($\beta = -0.32$, $p = 0.009$). Interestingly, participants who completed SPACE using the Anchor control rated it less attractive than those who played the game using the Tap & Swipe control (see Fig. 8).

5.4. Study 2: UI widget

In Study 2, we assessed differences in performance and self-reported usability when adding a novel UI element that provides feedback to players on the extent of the rotations they performed in the VE. Specifically, the widget consists of a circle positioned on the right side of the HUD that gets filled in according to the rotation angle completed by the player (Fig. 5b). Critically, this circle only gets updated when the player is standing still, limiting the amount of information that is provided to solve the task. The widget was introduced based on feedback we received in Study 1. For Study 2, we compared the Widget against the Anchor condition from Study 1 since the latter proved to be the most suitable control interface to interact with SPACE.

5.4.1. Methods

A power analysis for a linear multiple regression with four predictors revealed that a sample size of 129 participants was sufficient to detect a moderate effect size ($f^2 = 0.15$), with a power of 0.95 and α set at 0.05. In total, 128 participants were included in this study. Ninety-two participants consisted of those previously assigned to the Anchor condition in Study 1 (no widget), and 36 new participants were assigned to the widget condition (Widget). Sample characteristics for both conditions are the same despite the differences in the number of participants (see Table 1). The materials and procedure were performed as described in the general Methods section above, with the addition of the NASA-TLX

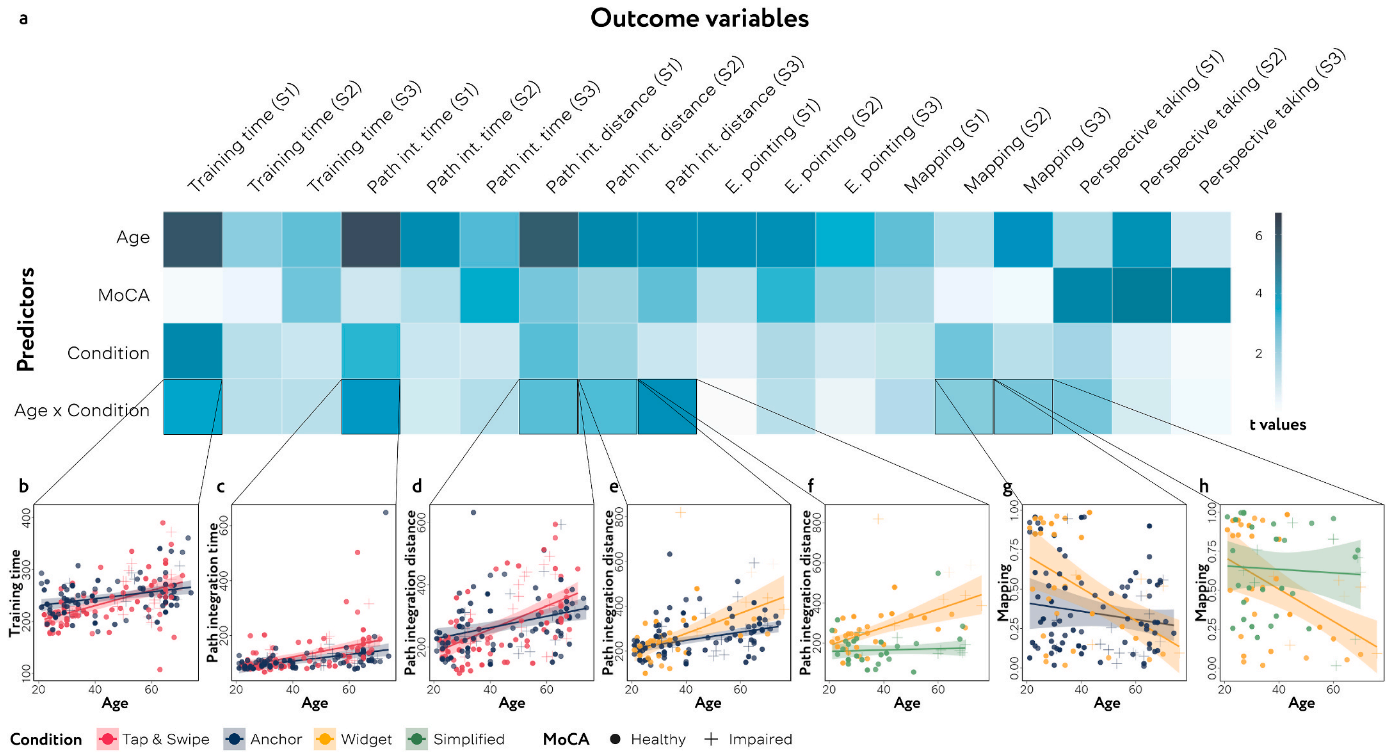


Fig. 7. Regression results with task performance as the outcome variables for the three studies (S1, S2 and S3). The coloured matrix presents the *t-values* for the combination of predictor and outcome variables for each model. The graphs below present the significant interactions between Age and Condition. Circles represent participants with MoCA scores above 25, and crosses represent participants with MoCA scores of 25 and below.

and Presence questionnaires.

5.4.2. Results

Descriptive statistics for the SPACE tasks and self-report usability questionnaires are presented in Table 3. Results from the regression models are presented in Figs. 7 and 8, and Table S2 in the Supplementary material. In terms of task performance, results revealed that *Age* significantly predicted training time ($\beta = 0.22$, $p = 0.049$), path integration time ($\beta = 0.17$, $p < 0.001$), path integration distance error ($\beta = 0.25$, $p < 0.001$), egocentric pointing error ($\beta = 0.33$, $p < 0.001$), and perspective taking error ($\beta = 0.35$, $p < 0.001$). Similar to Study 1, task performance decreased with age. In addition, *Cognitive status* significantly predicted errors in the egocentric pointing ($\beta = -0.23$, $p = 0.006$) and perspective taking ($\beta = -0.41$, $p < 0.001$) tasks. Here again, lower MoCA scores were associated with worse performance. *Condition* was also a significant predictor of mapping score ($\beta = 0.40$, $p = 0.024$). Specifically, accuracy in the mapping task was higher for participants who completed the previous navigation tasks with the help of the rotation widget. Finally, the interaction between *Age* and *Condition* also significantly predicted path integration distance error ($\beta = 0.35$, $p = 0.010$) and mapping score ($\beta = -0.43$, $p = 0.039$). While there was an overall increase in performance for both young and old participants, the presence of the widget was more advantageous for the younger participants.

Results from the self-reported usability questionnaires revealed that *Age* significantly predicted scores in the SUS ($\beta = -0.41$, $p < 0.001$), UEQ Pragmatic ($\beta = -0.38$, $p < 0.001$), NASA-TLX ($\beta = 0.27$, $p = 0.014$), and Presence questionnaire ($\beta = -0.28$, $p = 0.007$). Specifically, self-reported usability and presence decreased with *Age*, while perceived workload increased with *Age*. There were no other significant effects.

5.5. Study 3: configuration

In Study 3, we assessed differences in performance and self-reported usability when using a simplified configuration of landmarks in the VE in order to identify a more concise and user-friendly version of the game (Fig. 5c). Specifically, we reduced the number of landmarks to three and the number of trials to seven in the Simplified condition, as opposed to the five landmarks and thirteen trials in the other conditions. Additionally, we decreased the area of the polygon occupied by the landmarks in the Simplified condition to 4.5, compared to 27 in the other conditions. These changes were motivated by previous work that found that task complexity is a critical feature for the acceptance of technology by older adults (Cota & Ishitani, 2015; Czaja & Lee, 2012; Flores et al., 2008; Gerling et al., 2012; Hamid et al., 2023; Moxley et al., 2022). Indeed, several participants in Studies 1 and 2 observed that the game was too demanding and lengthy.

5.5.1. Methods

A power analysis for a linear multiple regression with four predictors revealed that a sample size of 67 participants was sufficient to detect a medium to large effect size ($f^2 = 0.30$), with a power of 0.95 and α set at 0.05. In total, 72 participants were included in this study. Thirty-six participants consisted of those previously assigned to the Widget condition, and 36 new participants were assigned to the Simplified condition. The materials and procedure were performed as described in the general Methods section with the addition of the NASA-TLX and Presence questionnaires.

5.5.2. Results

Descriptive statistics for the SPACE tasks and self-report usability questionnaires are presented in Table 3. Results from the regression models are presented in Figs. 7 and 8, and Table S3 in the Supplementary material. In terms of task performance, results revealed that *Age* significantly predicted training time ($\beta = 0.56$, $p = 0.016$), path integration time ($\beta = 0.31$, $p = 0.010$), path integration distance error ($\beta = 0.40$, $p < 0.001$), pointing error ($\beta = 0.46$, $p = 0.004$), and mapping score ($\beta = -0.53$, $p = 0.001$). Similar to Studies 1 and 2, task

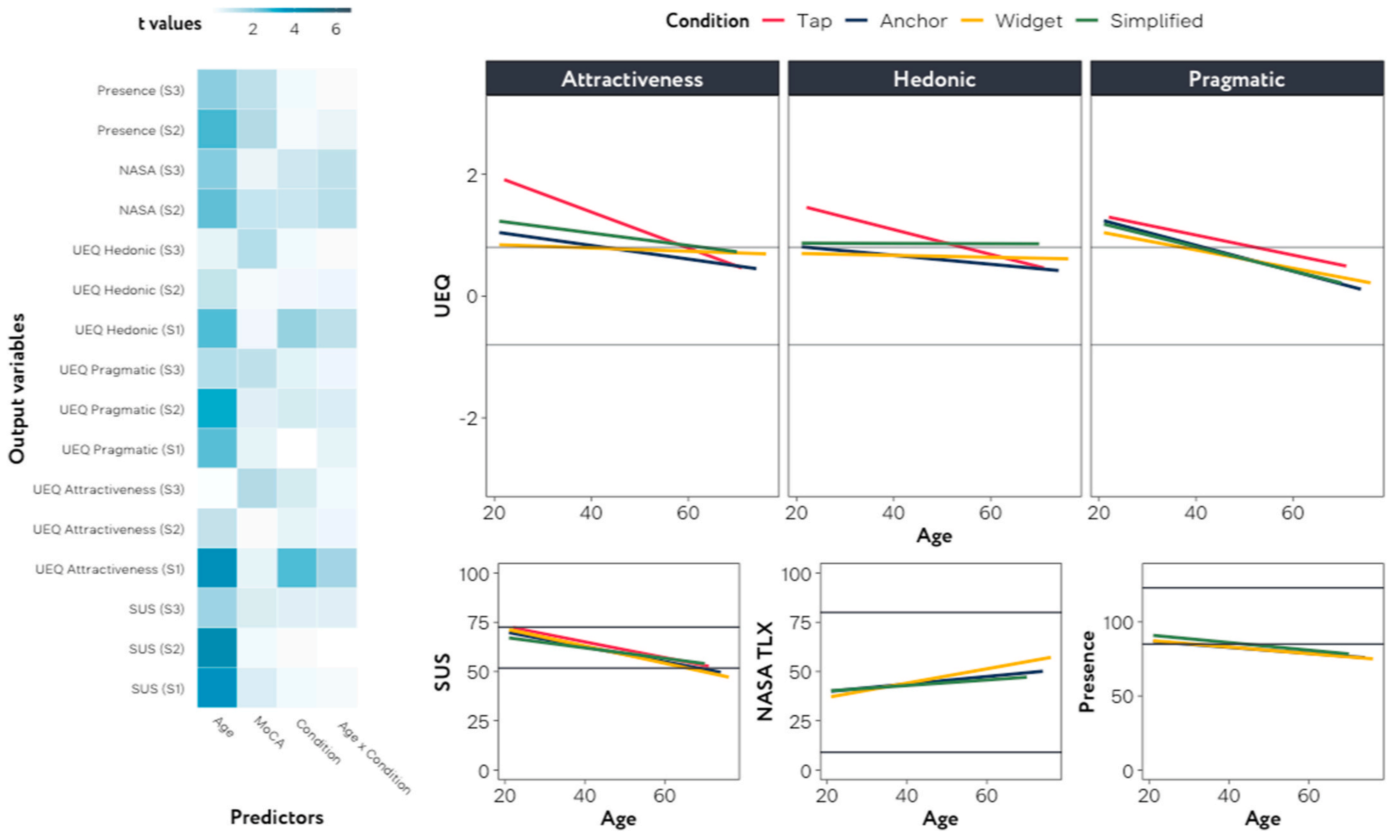


Fig. 8. Regression results with the self-reported usability questionnaires as outcome variables for the three studies (S1, S2 and S3). The coloured matrix presents the *t-values* for the combination of predictor and outcome variables for each model. The graphs present the usability questionnaire results relative to age for the four experimental conditions. The black lines in the graphs indicate the benchmarked accepted values for each questionnaire. UEQ scores range from "Neutral" (−0.8) to "Positive" (0.8), SUS range from "OK" (51.7) to "Good" (72.5), NASA-TLX range from "Medium" (10) to "High" (79), and Presence range from "Below" (85) to "Above" (123) average.

performance decreased with age. In addition, *Cognitive status* significantly predicted error scores in training time ($\beta = -0.26$, $p = 0.023$), path integration time ($\beta = -0.27$, $p = 0.002$), path integration distance ($\beta = -0.25$, $p = 0.016$), and perspective taking ($\beta = -0.61$, $p < 0.001$). Here, lower MoCA scores were associated with worse performance. *Condition* did not predict the performance of any of the tasks. Finally, the interaction between *Age* and *Condition* significantly predicted path integration distance error ($\beta = -0.48$, $p < 0.001$) and mapping score ($\beta = 0.47$, $p = 0.031$). In both tasks, performance decreased with age for participants in the complex configuration group but not for those in the simplified configuration group.

Results from the self-reported usability questionnaires revealed that *Age* significantly predicted scores in the NASA-TLX ($\beta = 0.38$, $p = 0.046$), with the self-reported workload increasing with age. There were no other significant effects.

5.6. Qualitative analysis

5.6.1. Results

We collected a total of 836 unique responses from 245 (95.7%) participants, of whom 49% belonged to the Young group, 23% to the Middle-aged group, and 28% to the Old group. Our thematic analysis unveiled six major themes (i.e., Enjoyable, Intuitive, Stimulating, Demanding, Repetitive, and Frustrating), each reflecting distinct attributes experienced by the participants during gameplay. Fig. 9 presents a Sankey diagram that visualises the frequency of responses for each condition with respect to the different themes separated by age group.

5.6.1.1. Enjoyable. Over one quarter (27%) of the responses were related to the enjoyment, satisfaction and visual appeal of the game. These comments highlight the positive attitudes that participants had towards the gameplay experience and the overall pleasantness of specific in-game tasks (e.g., "*[I liked] the mystery to be uncovered in locating the rocket*"). Some of these participants enjoyed the aesthetic appeal of the graphics (e.g., "*The scenery and objects are beautiful and can [re]present the simulation really well*") and the atmosphere created by the soundtrack (e.g., *"Playful character and fun music to play the games*"). In terms of enjoyment, a higher number of responses from Old participants (42%) indicated that playing SPACE in the simplified condition was more enjoyable than playing SPACE in the Tap & Swipe (25%), Anchor (25%), and Widget (14%) conditions. Similarly, a higher proportion of responses from Middle-aged participants indicated that the Simplified condition was more enjoyable (38%) compared to Tap & Swipe (25%), Anchor (24%), and Widget conditions (24%). A higher proportion of responses from Young participants indicated that playing SPACE in the Tap & Swipe condition was more enjoyable (33%) compared to the Anchor (28%), Widget (27%), and Simplified (28%) conditions. We believe that this difference is likely due to the freedom these participants experienced with the control interface in the Tap & Swipe condition.

5.6.1.2. Intuitive. Approximately one-quarter of the responses (24%) were related to the intuitiveness and ease-of-use of the game's UI (e.g., "*Visually and functionally simple. Easy to use*"), controls (e.g., "*Controls were easy to use*") and the ease of finding enough information to perform the tasks (e.g., "*Easy to follow. Simple and minimal steps/procedure*"). In terms of intuitiveness, a higher proportion of responses from Old participants indicated that playing SPACE with the Simplified condition (27%) was more intuitive compared to the Widget (24%), Anchor

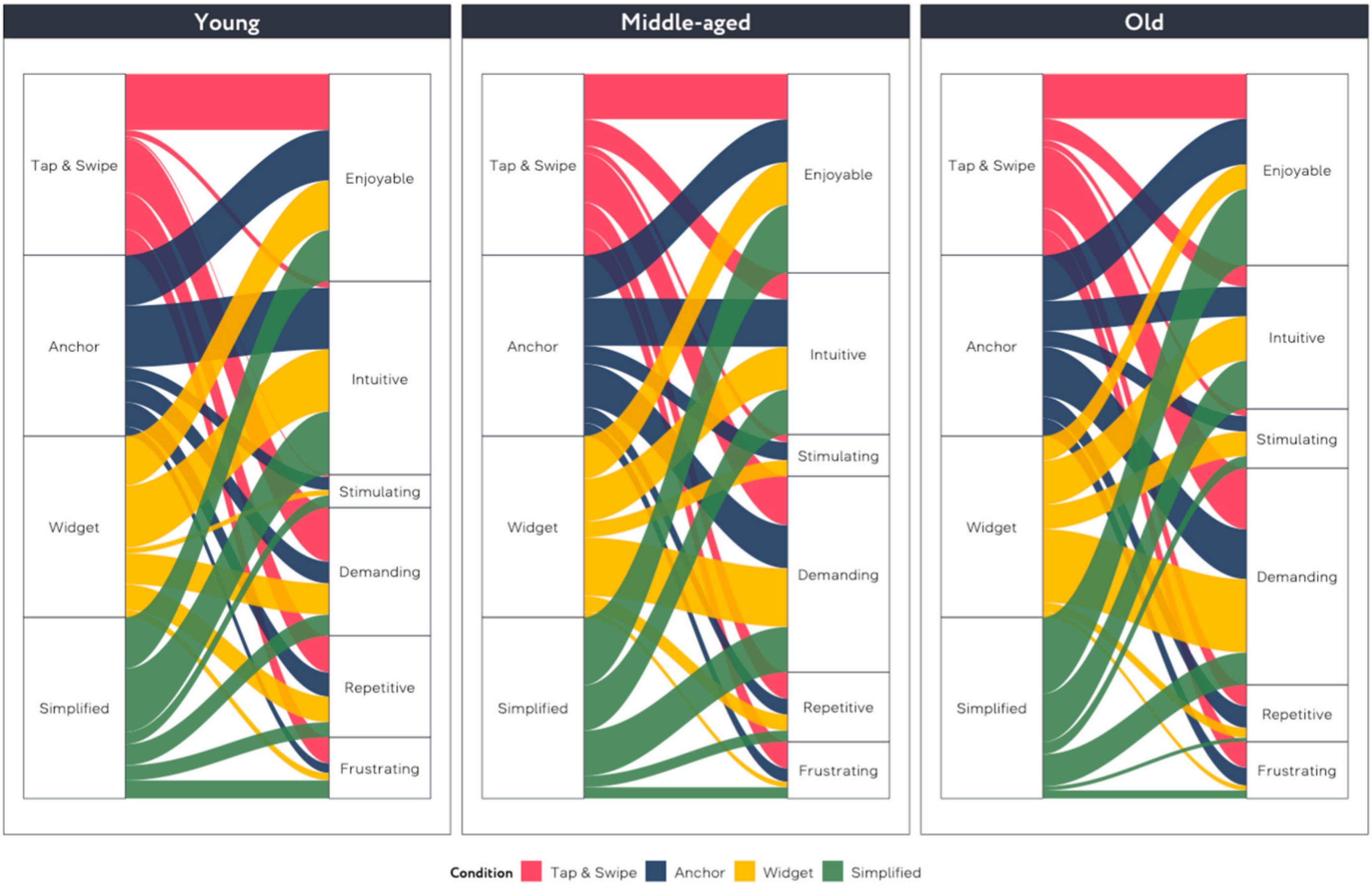


Fig. 9. A Sankey diagram visualising the flow frequency between the four experimental conditions and the six themes for the three age groups.

(16%), and Tap & Swipe (12%) conditions. A similar upward trend from the initial Tap & Swipe condition is also present for both the Young (Tap = 4%; Anchor = 34; Widget = 34%; Simplified = 35%) and Middle-aged (Tap = 15%; Anchor = 26%; Widget = 24%; Simplified = 25%) participants.

5.6.1.3. Stimulating. Only a small number of responses (6%) were related to the game's capacity to stimulate and captivate interest. Here, players highlighted the engaging nature of the game (e.g., "*Path integration was addictive*") and its potential as a cognitive training tool (e.g., "*This game is creative and helps to train the spatial and cognitive awareness of the gamers. Cool game*"). Results revealed that a higher proportion of responses from Old participants indicated that playing SPACE in the Widget (14%) and Anchor (9%) conditions was more stimulating compared to the Simplified (7%) and Tap & Swipe (4%) conditions. A similar trend is also present for the Middle-aged group (Widget = 9%; Anchor = 10%; Tap = 4%; Simplified = NA). In contrast, the higher proportion of responses from Young participants indicated that playing SPACE in the Simplified (7%) and Anchor (7%) conditions was more stimulating compared to the Widget (3%) and Tap & Swipe (1%) conditions.

5.6.1.4. Demanding. Approximately 23% of responses highlighted that the game was demanding. Interestingly, most of the comments on this theme were related to challenges in solving the tasks, particularly path integration (e.g., "*The task was complicated. Although the instructions were clear, it was still difficult to navigate*"). Participants also remarked that it was difficult to estimate distances and rotations during gameplay (e.g., "*It was hard to find the object's distance and angle*") and that the path integration task seemed unrealistic (e.g., "*Rocket task was difficult as there were no landmarks to guide me back, the background looks the same for all*"). In terms of task demand, a higher proportion of responses from Old participants (34%) indicated that playing SPACE with the Tap & Swipe condition was more demanding, compared to the Anchor (25%), Widget (41%) and Simplified (18%) conditions. A similar downward trend is present for both the Young (Tap = 30%; Anchor = 12; Widget 17%; Simplified = 12%) and Middle-aged (Tap = 27%; Anchor = 24%; Widget = 32%; Simplified = 25%) participants.

5.6.1.5. Repetitive. Only 12% of the responses emphasised that some aspects of the game were repetitive. Here, participants noted that some tasks took too long to complete (e.g., "*The game is a bit too long. [I will] get bored and not motivated anymore*"), had too many repetitive trials (e.g., "*A little repetitive*"), and were boring (e.g., "*Too slow and long, very boring*"). In terms of repetitiveness, the proportion of responses was consistent across all groups with younger participants reporting overall higher percentages. Specifically, these groups found that playing SPACE in the Simplified condition was less repetitive (Old = 2%; Middle-aged = 6%; Young = 8%) compared to the Widget (Old = 9%; Middle-aged = 9%; Young = 14%), Anchor (Old = 12%; Middle-aged = 9%; Young = 13%) and Tap & Swipe (Old = 12%; Middle-aged = 15%; Young = 20%) conditions.

5.6.1.6. Frustrating. A very small number (8%) of the responses indicated that some participants became frustrated while playing the game. Players expressed their frustration by noting that they became discouraged by their performance on some of the tasks (e.g., "*I wanted to stop playing after getting lost*") or because they struggled to interact with the controls of the game (e.g., "*I was struggling [with] the touch and swipe controls*"). In terms of frustration, a lower proportion of responses from Old and Middle-aged participants indicated that playing SPACE in the Widget (Old = 3%; Middle-aged = 3%) and Simplified conditions (Old

= 4%; Middle-aged = 6%) was less frustrating compared to the Anchor (Old = 10%; Middle-aged = 8%) and Tap & Swipe (Old = 14%; Middle-Aged = 14%) conditions. A higher proportion of responses from Young participants indicated that playing SPACE in the Tap & Swipe (14%) and Simplified (14%) conditions was more frustrating compared to the Anchor (4%) and Widget (5%) conditions. Altogether, these results indicate that the Tap condition was equally frustrating for all groups. In addition, Young participants may have found the Simplified condition more frustrating because it was not sufficiently challenging.

6. Discussion

In three studies, we investigated the usability of a novel iPad game designed to detect early signs of cognitive impairment. In Study 1, we designed a semiautomated control interface (Anchor) that required participants only to tap and hold a button to navigate and compared it to the traditional Tap & Swipe interface used in games. In Study 2, we introduced a UI widget to the HUD that informed players about the amount of rotation they performed in the VE and assessed its benefits relative to playing the game without it. In Study 3, we created a simplified configuration of trials and landmarks and evaluated its effect compared to the more complex configuration used in the previous studies.

Across all studies, the age of the participants was the strongest predictor of task performance. In general, younger participants outperformed older participants in the training (time), path integration (time and distance error), egocentric pointing (angle error), mapping (r^2) and perspective taking (angle error) tasks. As expected, age did not always predict performance in the perspective taking task (Studies 1 and 2), suggesting a dissociation between this task and the other tasks in SPACE. Indeed, success on the perspective taking task depends on spatial knowledge rather than information acquired during the previous tasks in SPACE. Compared to older participants, younger participants also consistently reported higher usability (Studies 1, 2, and 3), lower perceived workload (Studies 2 and 3) and a higher sense of presence (Study 2) after playing SPACE. This difference in performance and usability ratings across ages may stem from older adults facing greater motor, cognitive, and perceptual challenges with digital technologies (Barnard et al., 2013; Czaja et al., 2019; Gamberini et al., 2006; Gerling et al., 2012; Martinho et al., 2020), leading to diminished confidence (Czaja & Lee, 2012), motivation (Czaja et al., 2019), and familiarity with these tools (Holthe et al., 2018; Kaufman et al., 2016). While all participants in the studies successfully completed a thorough training phase before the spatial tasks, it is still possible that these factors prevented them from performing at the same level as younger participants. This result is consistent with previous research that found age differences for various digital instruments designed to assess spatial and navigation abilities (Coutrot et al., 2018; Howett et al., 2019; Stangl et al., 2020; Tu et al., 2015; van der Ham et al., 2020).

The manipulations we introduced with regard to the control interface, UI widget and configuration complexity only had a minimal effect on perceived usability. This result was unexpected, given the large performance gains across conditions observed for all tasks in SPACE. Despite decreasing with age, the usability ratings for the UEQ and, to a lesser extent, the SUS remained within the accepted levels for these questionnaires across all conditions. For the UEQ, scores for the attractiveness, hedonic and pragmatic subcategories revealed that older participants had neither a negative nor a positive user experience playing SPACE. In other words, older adults found the technology functional and useable but not particularly engaging. In contrast, the younger participants had a more positive evaluation of SPACE, suggesting that they perceived the technology favourably in its appeal, ease of use, and overall significance. In relation to the UEQ benchmark devised by Schrepp et al. (2017), younger participants rated SPACE above the level of mature products designed for commercial purposes. In contrast, the neutral evaluations by the older participants (+40 y/o UEQ Pragmatic; +50 y/o UEQ Attractiveness and Hedonic) suggest that additional adjustments may be necessary to optimise their overall experience. Based on the criteria established by Bangor et al. (2009), the SUS scores for older participants were situated between High and Low acceptability standards and have a level of usability that can be defined as “OK”. Notably, SUS scores were higher for the Simplified condition, highlighting that lower levels of complexity resulted in the highest usability gains for these participants. In contrast, SUS scores for the younger participants were situated in the High acceptability zone, ranging between “OK” to “Good”. Interestingly, the usability scores for younger participants decreased with each study manipulation, suggesting that lower complexity may have led to less engagement and user satisfaction for these participants.

Age was also a significant predictor of perceived workload in Studies 2 and 3. For both these studies, younger participants reported lower cognitive load than older participants. Critically, the perceived workload load scores for younger and older participants were only slightly above the median threshold of 41, as established by Hertzum (2021) for the task load index of VR products. While the scores for younger participants from the NASA-TLX did not vary across conditions, older participants reported lower cognitive load in the Simplified condition. The Presence scores for the younger and older participants were below the benchmarked value (M = 104.38, SD = 18.99) proposed by the Laboratoire de Cyberpsychologie de l’UQO (Witmer et al., 2005). As expected, older participants consistently reported lower scores compared to younger participants, and these differences were significant after the introduction of the rotation widget. Given the large standard deviation in benchmark scores, results from the three studies in SPACE suggest that some improvements can be made but that the game is already sufficiently immersive for a tablet app. Overall, the significant correlations between the questionnaires across all studies suggest that usability, perceived workload, and presence are inherently interconnected, where increased usability and presence coincide with decreased workload.

In all studies, the significant interaction between age and condition suggests that younger and older participants benefited differently from the experimental manipulations. In Study 1, older participants using the Anchor control showed greater improvement in the training and path integration tasks than younger participants. Sensorimotor skills are known to play a critical role in gaming by allowing participants to perform complex interactions without losing focus of the goals of the game (Flores et al., 2008; Gerling et al., 2012; Ijsselsteijn et al., 2007). Here, the semi-automated and user-friendly design of the Anchor control may have facilitated the interaction of older participants with the controls by reducing the challenges associated with the precise finger movements required in the Tap & Swipe interface. The Anchor control interface breaks down navigation in the VE into smaller manageable steps by separating the rotation and translation components of locomotion. This approach makes the task instructions easier to understand and reduces the information load given to the participant. Consequently, participants can optimise their cognitive resources by focusing on completing one task at a time. These findings align with studies suggesting that simplified interactions that are easy to understand are recommended when designing games for ageing adults (Gamberini et al., 2006; Gerling et al., 2012; Marston, 2013). Interestingly, younger participants performed better with the Tap & Swipe than with the Anchor control. Here, the fragmented nature of the Anchor control compared to the canonical gestures used in Tap & Swipe control may have restricted the freedom of their interactions with the VE and affected their performance.

In Study 2, we found that certain UI elements can be beneficial for the performance of younger and older participants, but the extent of these benefits depends on the task. The addition of the rotation widget led to better outcomes in the path integration tasks among younger participants, whereas it resulted in poorer performance for older participants. The presence of the widget also led to better performance from younger and older participants in the mapping task, although younger participants benefited more from its feedback than older participants. While previous research suggests that on-screen aids (e.g., maps, arrows) are generally beneficial for younger and older adults (Czaja & Lee, 2012; Gramopadhye et al., 2014; Johanson et al., 2023), the current study reveals that these additions have a limited effect and, in some cases, can be detrimental to performance. As indicated by the NASA-TLX scores, the increased workload for older participants who completed the widget condition suggests that this navigation aid may have been challenging to use during the path integration task. However, the presence of rotation feedback appears to have assisted participants in developing configurational knowledge of the VE. This is evident as participants below the age of 60 demonstrated superior performance in the mapping task compared to individuals of the same age using the Anchor control. These findings underscore the need for careful consideration in the design and implementation of UI elements and challenge the assumption that all screen aids may be beneficial for older adults. Screen aids and widgets may be helpful, but they need to be sufficiently simple to not interfere with the task and overload the player with information (Gamberini et al., 2006).

In Study 3, we found that the Simplified configuration of landmarks and trials further mitigated age-related performance differences for some tasks. Notably, performance in the path integration and mapping tasks remained stable across all age groups for the Simplified configuration but declined with age for participants who played with the complex default configuration. Interestingly, the effect of the condition itself did not significantly affect task performance. Consistent with the literature, these results suggest that moderate simplifications in gameplay may effectively reduce age-related differences (Cota & Ishitani, 2015; Czaja & Lee, 2012; Flores et al., 2008; Gerling et al., 2012; Moxley et al., 2022) in performance that can sometimes mask actual ability. Here, the large variation in performance across all participants in these tasks suggests that despite the simplification, the tasks maintained an appropriate level of difficulty necessary for picking up differences related to cognitive status. Indeed, improved performance by older participants was associated with increased self-reported workload, but age differences were no longer evident in the SUS, UEQ subscales, and Presence scores.

Creating easily accessible serious games for testing cognitive status on a global scale is a challenging endeavour. While games like Sea Hero Quest (Coutrot et al., 2018) were designed to identify participants exhibiting signs of MCI, the oversimplification of some of the tasks likely led to a loss of sensitivity for detecting changes in cognitive status. Here, researchers face the dilemma of releasing unsupervised games with simpler tasks that are easier to play or burdening players in more controlled environments (e.g., labs and clinics) with challenging tasks that reduce engagement but are more suitable for identifying cognitive impairment. In the three studies, we used MoCA scores as a proxy of cognitive impairment to help us understand the extent to which performance in SPACE can be used as an indicator of cognitive status. While the MoCA cannot provide the same sensitivity as a full neuropsychological assessment, it is still the preferred, most accurate mode of assessment of cognitive status (Jia et al., 2021; Pinto et al., 2019; Tsai et al., 2016). Across all studies, we found that lower MoCA scores were associated with lower performance in the training, path integration, pointing, and perspective taking tasks. These findings collectively underscore the link between diminished cognitive abilities and the challenges faced in completing spatial tasks (Benke et al., 2014; Castegnaro et al., 2023; Coughlan et al., 2018; Cushman et al., 2008; deIpolyi et al., 2007; Hort et al., 2007; Howett et al., 2019; Segen et al., 2022). Remarkably, MoCA scores consistently predicted scores in the path integration (Studies 1 and 3) and the perspective taking (all studies) tasks, suggesting the potential of these SPACE tasks to contribute to future cognitive assessments.

The results from the qualitative analysis support the quantitative findings and show the overall advantages of the Simplified condition compared to the other conditions for participants of all ages. Specifically, participants from the Middle-aged and Old age groups consistently reported the Simplified condition as the (or among the) most enjoyable, more intuitive, less demanding, less repetitive, and less frustrating condition. This preference aligns with cognitive theories suggesting that easy and short tasks are preferable for older individuals (Gamberini et al., 2006; Gerling et al., 2012; Marston, 2013). Interestingly, Young participants found the Simplified condition as frustrating as the Tap & Swipe condition and reported greater enjoyment when playing SPACE with the Tap & Swipe control. Given their greater digital fluency and confidence using touchscreens, Young participants likely appreciated the increased freedom offered by the Tap & Swipe control interaction. Despite these remarks, the Simplified condition still delivered a notably positive gaming experience for Young participants, who found it more stimulating than the other conditions. Indeed, all three age groups highlighted the greater intuitiveness and reduced demands of the Simplified condition relative to the other conditions. In summary, the Simplified version of SPACE emerged as the most suitable configuration because it simultaneously enhanced enjoyment among Middle-aged and Old participants and was regarded as more intuitive, less repetitive, and demanding across all age groups.

There are a series of limitations associated with the three studies. First, the size of the sample was different across conditions and unbalanced in Study 2. To address this limitation, we computed separate power analyses to ensure that the effect sizes for each study were meaningful, employed robust statistical methods to account for potential heteroscedasticity in the data, and ensured a proportional representation of age groups across conditions. Second, only a small number of participants aged 70 and above were recruited. While we acknowledge this limitation in our sampling it is important to highlight that SPACE was primarily designed for the early detection of cognitive impairment. Third, the predominance of Singaporean participants in our studies can influence the generalizability of the findings, as some results were likely shaped by known country and cultural differences in spatial ability, education, and health (Coutrot et al., 2018; Ishikawa & Newcombe, 2021). Notwithstanding this possibility, our analyses of the sample characteristics revealed no significant differences between the four experimental conditions. Fourth, we based the usability changes on the iterative feedback that we received at the end of each study. Among this feedback, participants made other less prominent remarks that were not acted upon. For example, some participants inquired about the possibility of playing the game from a third-person perspective or being able to locomote backwards. While the choice of study manipulation was based on the incidence of the feedback, it is still possible that other usability issues still need to be addressed in SPACE. Fifth, our measure of cognitive impairment was based on the MoCA, which is known to be highly sensitive in discriminating between healthy and severely impaired individuals but is not diagnostic of pathology. As such, the link between performance in SPACE and cognitive impairment in future studies will require additional ground truthing.

Future research will be essential to determine whether SPACE can live up to its potential in identifying early signs of cognitive impairment. Towards this end, it will be essential to establish age and gender-based performance benchmarks for each task in SPACE to differentiate cognitive deficits from natural variations in spatial ability. These benchmarks will need to be calibrated relative to full neuropsychological assessments for a more accurate diagnosis of cognitive impairment. The potential of SPACE as an early digital marker for cognitive impairment will also depend on longitudinal cohort studies complemented by fluid and imaging biomarkers of neurodegeneration. SPACE was originally developed for controlled data collection in laboratories and clinics under supervised administration. The results of this usability study will lay the groundwork for further refinement, paving the way for mass deployment in unsupervised settings.

7. Conclusions

This paper presents a comprehensive investigation into the usability of an innovative iPad game designed to detect early signs of cognitive impairment. Across three studies, we found that older adults faced challenges with the standard tap and swipe control interface, the cognitive load imposed an additional UI element, and the general complexity resulting from the default configuration of trials and landmarks. The introduction of the semi-automated control interface and the simplified configuration of trials and landmarks resulted in significant improvements in performance for older adults. Interestingly, younger participants did not benefit from the semi-automated control interface and tended to perceive it as limiting. In contrast, younger participants were able to take advantage of the rotation widget and use it to fulfil the demands of the task and develop a configurational knowledge of the VE. Notably, the simplified condition that combined the semi-automated control interface, rotation widget, and less complex configuration proved to be the most suitable for both younger and older participants. Here, differences in performance in the Simplified condition were primarily associated with the overarching goal of SPACE in detecting cognitive impairment rather than age. A serious game aimed at detecting early signs of cognitive impairment must be sufficiently adaptable to accommodate the varying needs of different age groups. The results of our studies showed that it is essential to consider both task performance and self-reported usability in order to simultaneously capture the subjective satisfaction of the users and the efficiency of their interaction.

Data supporting this article may be available upon request and pending approval from the ethics committees in Singapore and ETH Zurich.

Funding statement

This research is supported by the National Research Foundation Singapore (NRF) under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

CRediT authorship contribution statement

Giorgio Colombo: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Karolina Minta:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jascha Grübel:** Writing – review & editing, Software, Data curation. **Wei Lin Eunice Tai:** Investigation, Formal analysis, Data curation. **Christoph Hölscher:** Writing – review & editing. **Victor R. Schinazi:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data supporting this article may be available upon request and pending approval from the ethics committees in Singapore and ETH Zurich.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2024.108349.

References

Abbott, A. (2013). Gaming improves multitasking skills. *Nature, 501*(7465), 18–18.

Adcock, M., Sonder, F., Schättin, A., Gennaro, F., & de Bruin, E. D. (2020). A usability study of a multicomponent video game-based training for older adults. *European Review of Aging and Physical Activity: Official Journal of the European Group for Research into Elderly and Physical Activity, 17*(1).

Alahäivälä, T., & Oinas-Kukkonen, H. (2016). Understanding persuasion contexts in health gamification: A systematic analysis of gamified health behavior change support systems literature. *International Journal of Medical Informatics, 96*, 62–70.

Amagai, S., Pila, S., Kaat, A. J., Nowinski, C. J., & Gershon, R. C. (2022). Challenges in participant engagement and retention using mobile health apps: Literature review. *Journal of Medical Internet Research, 24*(4), Article e35120.

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., Kong, E., Larraburo, Y., Rolle, C., Johnston, E., & Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. *Nature, 501*(7465), 97–101.

Arevalo-Rodriguez, I., Smailagic, N., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., Roqué i Figuls, M., Pedraza, O. L., Bonfill Cosp, X., & Cullum, S. (2013). Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). In *Cochrane database of systematic reviews*. John Wiley & Sons, Ltd.

Assila, A., Oliveira, K., & Ezzedine, H. (2016). Standardized usability questionnaires: Features and quality focus. *Computer Science and Information Technology, 6*.

Bahar-Fuchs, A., Clare, L., & Woods, B. (2013). Cognitive training and cognitive rehabilitation for mild to moderate Alzheimer's disease and vascular dementia. *Cochrane Database of Systematic Reviews, 2013*(6), Article CD003260.

Bamidis, P. D., Fissler, P., Papageorgiou, S. G., Zilidou, V., Konstantinidis, E. I., Billis, A. S., Romanopoulou, E., Karagianni, M., Beratis, I., Tsapanou, A., Tsilikopoulou, G., Grigoriadou, E., Ladas, A., Kyrillidou, A., Tsolaki, A., Frantzidis, C., Sidiropoulos, E., Siountas, A., Matsi, S., … Kolassa, I.-T. (2015). Gains in cognition through combined cognitive and physical training: The role of training dosage and severity of neurocognitive disorder. *Frontiers in Aging Neuroscience, 7*, 152.

Bang, M., Jang, C. W., Kim, H. S., Park, J. H., & Cho, H. E. (2023). Mobile applications for cognitive training: Content analysis and quality review. *Internet Interventions, 33*, Article 100632.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies, 4*(3), 114–123.

Baranowski, T., Buday, R., Thompson, D. I., & Baranowski, J. (2008). Playing for real: Video games and stories for health-related behavior change. *American Journal of Preventive Medicine, 34*(1), 74–82.

Barnard, Y., Bradley, M. D., Hodgson, F., & Lloyd, A. D. (2013). Learning to use new technologies by older adults: Perceived difficulties, experimentation behaviour and usability. *Computers in Human Behavior, 29*(4), 1715–1724.

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and Aging, 23*(4), 765–777.

Belchior, P., Yam, A., Thomas, K. R., Bavelier, D., Ball, K. K., Mann, W. C., & Marsiske, M. (2019). Computer and Videogame interventions for older adults' cognitive and everyday functioning. *Games for Health Journal, 8*(2), 129–143.

Ben-Sadoun, G., Sacco, G., Manera, V., Bourgeois, J., König, A., Foulon, P., Fosty, B., Bremond, F., d'Arripe-Longueville, F., & Robert, P. (2016). Physical and cognitive stimulation using an Exergame in Subjects with normal aging, mild and moderate cognitive impairment. *Journal of Alzheimer's Disease: JAD, 53*(4), 1299–1314.

Benke, T., Karner, E., Petermichl, S., Prantner, V., & Kemmler, G. (2014). Neuropsychological deficits associated with route learning in Alzheimer disease, MCI, and normal aging. *Alzheimer Disease and Associated Disorders, 28*(2), 162–167.

Berg, J.-L., Durant, J., Léger, G. C., Cummings, J. L., Nasreddine, Z., & Miller, J. B. (2018). Comparing the electronic and standard versions of the Montreal cognitive assessment in an outpatient memory disorders clinic: A validation study. *Journal of Alzheimer's Disease: JAD, 62*(1), 93–97.

Berron, D., Glanz, W., Clark, L., Basche, K., Grande, X., Güsten, J., Billette, O. V., Hempen, I., Naveed, M. H., Diersch, N., Butryn, M., Spottke, A., Buerger, K., Perneczky, R., Schneider, A., Teipel, S., Wiltfang, J., Johnson, S., Wagner, M., … Düzel, E. (2024). A remote digital memory composite to detect cognitive impairment in memory clinic samples in unsupervised settings using mobile devices. *NPJ Digital Medicine, 7*(1), 79.

Blennow, K., Dubois, B., Fagan, A. M., Lewczuk, P., de Leon, M. J., & Hampel, H. (2015). Clinical utility of cerebrospinal fluid biomarkers in the diagnosis of early Alzheimer's disease. *Alzheimer's and Dementia: The Journal of the Alzheimer's Association, 11*(1), 58–69.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In *Usability evaluation in Industry* (pp. 207–212). CRC Press.

Castegnaro, A., Ji, Z., Rudzka, K., Chan, D., & Burgess, N. (2023). Overestimation in angular path integration precedes Alzheimer's dementia. *Current Biology: CB*.

Castillo Escamilla, J., León Estrada, I., Alcaraz-Iborra, M., & Cimadevilla Redondo, J. M. (2023). Aging: Working memory capacity and spatial strategies in a virtual orientation task. *GeroScience, 45*(1), 159–175.

Ceci, L. (2023). *Retention rate on day 30 of mobile app installs worldwide in 3rd quarter 2022, by category*. Statista. https://www.statista.com/statistics/259329/ios-and-android-app-user-retention-rate/.

Chan, D., Gallaher, L. M., Moodley, K., Minati, L., Burgess, N., & Hartley, T. (2016). The 4 Mountains test: A short test of spatial memory with high sensitivity for the diagnosis of pre-dementia Alzheimer's disease. *Journal of Visualized Experiments: JoVE, 116*.
Chételat, G., Villemagne, V. L., Bourgeat, P., Pike, K. E., Jones, G., Ames, D., Ellis, K. A., Szoeke, C., Martins, R. N., O'Keefe, G. J., Salvado, O., Masters, C. L., & Rowe, C. C. (2009). Relationship between atrophy and Î2-amyloid deposition in Alzheimer's disease. *Annals of Neurology*. NA – NA.
Chong, J. R., Ashton, N. J., Karikari, T. K., Tanaka, T., Schöll, M., Zetterberg, H., Blennow, K., Chen, C. P., & Lai, M. K. P. (2021). Blood-based high sensitivity measurements of beta-amyloid and phosphorylated tau as biomarkers of Alzheimer \textquoterights disease: A focused review on recent advances. *Journal of Neurology, Neurosurgery, and Psychiatry, 92*(11), 1231–1241.
Clement, J. (2023). ***Distribution of gamers in selected gaming markets in Europe in 2023, by age group***. Statista. https://www.statista.com/forecasts/448467/gaming-in-europe-by-age-and-country.
Contreras-Somoza, L. M. ia, Irazoki, E., Toribio-Guzmán, J. M., de la Torre-D' iez, I., Diaz-Baquero, A. A., Parra-Vidales, E., Perea-Bartolomé, M. ia V., & Franco-Mart' in, M.Á. (2021). Usability and user experience of cognitive intervention technologies for elderly people with MCI or dementia: A systematic review. *Frontiers in Psychology, 12*.
Cota, T. T., & Ishitani, L. (2015). Motivation and benefits of digital games for the elderly: A systematic literature review. *Revista Brasileira de Computação Aplicada, 7*(1).
Coughlan, G., Coutrot, A., Khondoker, M., Minihane, A.-M., Spiers, H., & Hornberger, M. (2019). Toward personalized cognitive diagnostics of at-genetic-risk Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America, 116*(19), 9285–9292.
Coughlan, G., Laczó, J., Hort, J., Minihane, A.-M., & Hornberger, M. (2018). Spatial navigation deficits - overlooked cognitive marker for preclinical Alzheimer disease? *Nature Reviews Neurology, 14*(8), 496–506.
Coutrot, A., Silva, R., Manley, E., de Cothi, W., Sami, S., Bohbot, V. D., Wiener, J. M., Hölscher, C., Dalton, R. C., Hornberger, M., & Spiers, H. J. (2018). Global Determinants of navigation ability. *Current Biology: CB, 28*(17), 2861–2866.e4.
Cushman, L. A., Stein, K., & Duffy, C. J. (2008). Detecting navigational deficits in cognitive aging and Alzheimer disease using virtual reality. *Neurology, 71*(12), 888–895.
Czaja, S. J. (2019). Usability of technology for older adults: Where are We and where Do We need to Be. *Usability Professionals' Association*.
Czaja, S. J., Beach, S., Charness, N., & Schulz, R. (2013). Older adults and the adoption of healthcare technology: Opportunities and challenges. In *Technologies for active aging* (pp. 27–46). Springer US.
Czaja, S. J., Boot, W. R., Charness, N., & Rogers, W. A. (2019). *Designing for older adults: Principles and creative Human factors Approaches* (3rd ed.). CRC Press.
Czaja, S. J., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., Rogers, W. A., & Sharit, J. (2006). Factors predicting the use of technology: Findings from the Center for research and education on aging and technology Enhancement (CREATE). *Psychology and Aging, 21*(2), 333–352.
Czaja, S. J., & Lee, C. C. (2012). Older adults and information technology: Opportunities and challenges. *Human Computer Interaction Handbook*, 825–840.
Dautzenberg, G., Lijmer, J., & Beekman, A. (2019). Diagnostic accuracy of the Montreal Cognitive Assessment (MoCA) for cognitive screening in old age psychiatry: Determining cutoff scores in clinical practice. Avoiding spectrum bias caused by healthy controls. *International Journal of Geriatric Psychiatry, 35*(3), 261–269.
deIpolyi, A. R., Rankin, K. P., Mucke, L., Miller, B. L., & Gorno-Tempini, M. L. (2007). Spatial cognition and the human navigation network in AD and MCI. *Neurology, 69*(10), 986–997.
Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification.". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning future media environments* (pp. 9–15).
Devanand, D. P., Pradhaban, G., Liu, X., Khandji, A., De Santi, S., Segal, S., Rusinek, H., Pelton, G. H., Honig, L. S., Mayeux, R., Stern, Y., Tabert, M. H., & de Leon, M. J. (2007). Hippocampal and entorhinal atrophy in mild cognitive impairment: Prediction of Alzheimer disease. *Neurology, 68*(11), 828–836.
Dias, L. P. S., Barbosa, J. L. V., & Vianna, H. D. (2018). Gamification and serious games in depression care: A systematic mapping study. *Telematics and Informatics, 35*(1), 213–224.
Diersch, N., & Wolbers, T. (2019). The potential of virtual reality for spatial navigation research across the adult lifespan. *Journal of Experimental Biology, 222*(Pt Suppl 1).
Entertainment Software Association. (2023). 2023 Essential facts about the US video game industry. ***Entertainment Software Association***. https://www.theesa.com/wp-content/uploads/2023/07/ESA_2023_Essential_Facts_FINAL_07092023.pdf.
Fagerholt, E., & Lorentzon, M. (2009). *Beyond the HUD - user interfaces for increased player immersion in FPS games*. Chalmers University of Technology.
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.
Faverio, M. (2022). ***Share of those 65 and older who are tech users has grown in the past decade***. Pew Research Center. https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/.
Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy, 98*, 19–38.
Flores, E., Tobon, G., Cavallaro, E., Cavallaro, F. I., Perry, J. C., & Keller, T. (2008). Improving patient motivation in game development for motor deficit rehabilitation. In *Proceedings of the 2008 International Conference on advances in computer entertainment technology*.
Folstein, M. F. (1983). The Mini-mental state Examination. *Archives of General Psychiatry, 40*(7), Article 812.
Friedman, A., & Kohler, B. (2003). Bidimensional regression: Assessing the configural Similarity and accuracy of cognitive maps and other two-Dimensional data sets. *Psychological Methods, 8*(4), 468–491.
Gamberini, L., Raya, M. A., Barresi, G., Fabregat, M., Ibanez, F., & Prontu, L. (2006). Cognition, technology and games for the elderly: An introduction to ELDERGAMES Project. *PsychNology J, 4*(3), 285–308.
Gerling, K. M., & Masuch, M. (2011). When gaming is not suitable for everyone: Playtesting wii games with frail elderly. In *1st Workshop on Game accessibility*. https://www.academia.edu/download/6449626/1.pdf.
Gerling, K. M., Schulte, F. P., Smeddinck, J., & Masuch, M. (2012). Game design for older adults: Effects of age-related changes on structural elements of digital games. In *Lecture Notes in computer Science* (pp. 235–242). Springer Berlin Heidelberg.
Gilbert, N. (2020). How many active video gamers were there in 2020? There were 2.69 billion video game players worldwide in 2020. ***FinancesOnline. Com***. https://financesonline.com/number-of-gamers-worldwide/.
Gramopadhye, N., Madathil, K. C., Bertrand, J., & Blair, C. S. (2014). An investigation of the effectiveness of navigational aids in a virtual environment. In *Proceedings of the Human factors and Ergonomics society … Annual Meeting Human factors and Ergonomics society. Meeting* (pp. 1949–1953), 58(1).
Granic, I., Lobel, A., & Engels, R. C. M. E. (2014). The benefits of playing video games. *American Psychologist, 69*(1), 66–78.
Groppell, S., Soto-Ruiz, K. M., Flores, B., Dawkins, W., Smith, I., Eagleman, D. M., & Katz, Y. (2019). A rapid, mobile neurocognitive screening test to aid in identifying cognitive impairment and dementia (BrainCheck): Cohort study. *JMIR Aging, 2*(1), Article e12615.
Grübel, J. (2022). Computer-aided experimentation for human behaviour analysis. *ETH Zurich]*.
Grübel, J. (2023). The design, experiment, analyse, and reproduce principle for experimentation in virtual reality. *Frontiers in Virtual Reality, 4*.
Grübel, J., Thrash, T., Hölscher, C., & Schinazi, V. R. (2017). Evaluation of a conceptual framework for predicting navigation performance in virtual reality. *PLoS One, 12*(9), Article e0184682.
Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature, 436*(7052), 801–806.
Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. In *2014 47th Hawaii International Conference on system Sciences* (pp. 3025–3034).
Hamid, S., Faith, F., Jaafar, Z., Abdul Ghani, N., & Yusop, F. D. (2023). Emerging technology for healthy Lifestyle of the middle-age and elderly: A scoping review. *Iranian Journal of Public Health, 52*(2), 230–242.
Hart, S. G. (2006). Nasa-task load index (NASA-TLX); 20 Years later. In *Proceedings of the Human factors and Ergonomics society … Annual Meeting Human factors and Ergonomics society. Meeting* (pp. 904–908), 50(9).
Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in Psychology* (pp. 139–183). Elsevier.
Havukainen, M., Laine, T. H., Martikainen, T., & Sutinen, E. (2020). A case study on Co-designing digital Games with older adults and Children: Game elements, Assets, and challenges. *The Computer Games Journal, 9*(2), 163–188.
Heart, T., & Kalderon, E. (2013). Older adults: Are they ready to adopt health-related ICT? *International Journal of Medical Informatics, 82*(11), e209–e231.
Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence, 30*(5), 425–447.
Hello Games. (2016). In ***No Man's Sky***. https://www.nomanssky.com/.
Hertzum, M. (2021). Reference values and subscale patterns for the task load index (TLX): A meta-analytic review. *Ergonomics, 64*(7), 869–878.
Holthe, T., Halvorsrud, L., Karterud, D., Hoel, K.-A., & Lund, A. (2018). Usability and acceptability of technology for community-dwelling older adults with mild cognitive impairment and dementia: A systematic literature review. *Clinical Interventions in Aging, 13*, 863–886.
Hort, J., Laczó, J., Vyhnálek, M., Bojar, M., Bureš, J., & Vlček, K. (2007). Spatial navigation deficit in amnestic mild cognitive impairment. *Proceedings of the National Academy of Sciences, 104*(10), 4042–4047.
Howett, D., Castegnaro, A., Krzywicka, K., Hagman, J., Marchment, D., Henson, R., Rio, M., King, J. A., Burgess, N., & Chan, D. (2019). Differentiation of mild cognitive impairment using an entorhinal cortex-based test of virtual reality navigation. *Brain: Journal of Neurology, 142*(6), 1751–1766.
Ijsselsteijn, W., Nap, H. H., de Kort, Y., & Poels, K. (2007). Digital game design for elderly users. In *Proceedings of the 2007 Conference on future play* (pp. 17–22).
Ishikawa, T., & Newcombe, N. S. (2021). Why spatial is special in education, learning, and everyday activities. *Cognitive Research: Principles and Implications, 6*(1), 20.
Jack, C. R., Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., … Silverberg, N. (2018). NIAAA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's and Dementia: The Journal of the Alzheimer's Association, 14*(4), 535–562.
Jack, C. R., Jr., & Holtzman, D. M. (2013). Biomarker modeling of Alzheimer's disease. *Neuron, 80*(6), 1347–1358.
Jack, C. R., Jr., Wiste, H. J., Vemuri, P., Weigand, S. D., Senjem, M. L., Zeng, G., Bernstein, M. A., Gunter, J. L., Pankratz, V. S., Aisen, P. S., Weiner, M. W., Petersen, R. C., Shaw, L. M., Trojanowski, J. Q., Knopman, D. S., & Alzheimer's Disease Neuroimaging Initiative. (2010). Brain beta-amyloid measures and magnetic

resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. *Brain: Journal of Neurology, 133*(11), 3336–3348.
Jia, X., Wang, Z., Huang, F., Su, C., Du, W., Jiang, H., Wang, H., Wang, J., Wang, F., Su, W., Xiao, H., Wang, Y., & Zhang, B. (2021). A comparison of the Mini-mental state Examination (MMSE) with the Montreal cognitive assessment (MoCA) for mild cognitive impairment screening in Chinese middle-aged and older population: A cross-sectional study. *BMC Psychiatry, 21*(1), Article 485.
Johanson, C., Gutwin, C., & Mandryk, R. (2023). Trails, rails, and over-reliance: How navigation assistance affects route-finding and spatial learning in virtual environments. *International Journal of Human-Computer Studies, 178*, Article 103097.
Kakulla, B. (2023). *The 50-Plus gamer of Today and Tomorrow*. AARP. https://www.aarp.org/content/dam/aarp/research/surveys_statistics/technology/2023/2023-gamers-50-plus.doi10.26419-2Fres.00585.001.pdf.
Kappen, D. L., Mirza-Babaei, P., & Nacke, L. E. (2019). Older adults' physical activity and exergames: A systematic review. *International Journal of Human-Computer Interaction, 35*(2), 140–167.
Kaufman, D., Sauvé, L., Renaud, L., Sixsmith, A., & Mortenson, B. (2016). Older adults' digital gameplay: Patterns, benefits, and challenges. *Simulation & Gaming, 47*(4), 465–489.
Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management, 45*, 191–210.
Konstantinidis, E. I., Billis, A. S., Paraskevopoulos, I. T., & Bamidis, P. D. (2017). The interplay between IoT and serious games towards personalised healthcare. In *2017 9th International Conference on virtual worlds and Games for serious applications (VS-Games)* (pp. 249–252).
Latham, A. J., Patston, L. L. M., & Tippett, L. J. (2013). The virtual brain: 30 years of video-game play and cognitive abilities. *Frontiers in Psychology, 4*.
Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In *Lecture Notes in computer Science* (pp. 63–76). Springer Berlin Heidelberg.
Lester, A. W., Moffat, S. D., Wiener, J. M., Barnes, C. A., & Wolbers, T. (2017). The aging navigational system. *Neuron, 95*(5), 1019–1035.
Levine, T. F., Allison, S. L., Stojanovic, M., Fagan, A. M., Morris, J. C., & Head, D. (2020). Spatial navigation ability predicts progression of dementia symptomatology. *Alzheimer's and Dementia: The Journal of the Alzheimer's Association, 16*(3), 491–500.
Li, A. W. Y., & King, J. (2019). Spatial memory and navigation in ageing: A systematic review of MRI and fMRI studies in healthy participants. *Neuroscience & Biobehavioral Reviews, 103*, 33–49.
Lim, J., Zhan, A., Ko, J., Terzis, A., Szanton, S., & Gitlin, L. (2012). A closed-loop approach for improving the wellness of low-income elders at home using game consoles. *IEEE Communications Magazine, 50*(1), 44–51.
Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. *JMIR Serious Games, 4*(2), e11.
Machado, M. da C., Ferreira, R. L. R., & Ishitani, L. (2018). Heuristics and Recommendations for the design of mobile serious Games for older adults. *International Journal of Computer Games Technology, 2018*.
Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., Conceicao, E. L. T., & Anna di Palma, M. (2023). robustbase. *Basic Robust Statistics*. http://robustbase.r-forge.r-project.org/.
Maggio, M. G., De Bartolo, D., Calabrò, R. S., Ciancarelli, I., Cerasa, A., Tonin, P., Di Iulio, F., Paolucci, S., Antonucci, G., Morone, G., & Iosa, M. (2023). Computer-assisted cognitive rehabilitation in neurological patients: State-of-art and future perspectives. *Frontiers in Neurology, 14*, Article 1255319.
Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods, 52*(2), 464–488.
Manera, V., Petit, P.-D., Derreumaux, A., Orvieto, I., Romagnoli, M., Lyttle, G., David, R., & Robert, P. H. (2015). "Kitchen and cooking," a serious game for mild cognitive impairment and Alzheimer's disease: A pilot study. *Frontiers in Aging Neuroscience, 7*, 24.
Marston, H. R. (2013). Design Recommendations for digital Game design within an ageing society. *Educational Gerontology, 39*(2), 103–118.
Martinho, D., Carneiro, J., Corchado, J. M., & Marreiros, G. (2020). A systematic review of gamification techniques applied to elderly care. *Artificial Intelligence Review, 53*(7), 4863–4901.
Mathews, S. C., McShea, M. J., Hanley, C. L., Ravitz, A., Labrique, A. B., & Cohen, A. B. (2019). Digital health: A path to validation. *Npj Digital Medicine, 2*(1).
McCord, A., Cocks, B., Barreiros, A. R., & Bizo, L. A. (2020). Short video game play improves executive function in the oldest old living in residential care. *Computers in Human Behavior, 108*, Article 106337.
McLennan, S. N., Mathias, J. L., Brennan, L. C., & Stewart, S. (2011). Validity of the montreal cognitive assessment (MoCA) as a screening test for mild cognitive impairment (MCI) in a cardiovascular population. *Journal of Geriatric Psychiatry and Neurology, 24*(1), 33–38.
McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M.-B. (2006). Path integration and the neural basis of the "cognitive map.". *Nature Reviews Neuroscience, 7*(8), 663–678.
Meier, I. B., Buegler, M., Harms, R., Seixas, A., Çöltekin, A., & Tarnanas, I. (2021). Using a digital Neuro Signature to measure longitudinal individual-level change in Alzheimer's disease: The Altoida large cohort study. *Npj Digital Medicine, 4*(1).
Migo, E. M., O'Daly, O., Mitterschiffthaler, M., Antonova, E., Dawson, G. R., Dourish, C. T., Craig, K. J., Simmons, A., Wilcock, G. K., McCulloch, E., Jackson, S. H. D., Kopelman, M. D., Williams, S. C. R., & Morris, R. G. (2016). Investigating virtual reality navigation in amnestic mild cognitive impairment using fMRI. Neuropsychology, Development, and Cognition. Section B, Aging. *Neuropsychology and Cognition, 23*(2), 196–217.
Morganti, F., Stefanini, S., & Riva, G. (2013). From allo- to egocentric spatial ability in early Alzheimer's disease: A study with virtual reality spatial tasks. *Cognitive Neuroscience, 4*(3–4), 171–180.
Morris, J. C., Storandt, M., Miller, J. P., McKeel, D. W., Price, J. L., Rubin, E. H., & Berg, L. (2001). Mild cognitive impairment represents early-stage Alzheimer disease. *Archives of Neurology, 58*(3).
Moxley, J., Sharit, J., & Czaja, S. J. (2022). The factors influencing older adults' Decisions Surrounding adoption of technology: Quantitative experimental study. *JMIR Aging, 5*(4), Article e39890.
Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699.
Nouchi, R., Taki, Y., Takeuchi, H., Hashizume, H., Akitsuki, Y., Shigemune, Y., Sekiguchi, A., Kotozaki, Y., Tsukiura, T., Yomogida, Y., & Kawashima, R. (2012). Brain training game improves executive functions and processing speed in the elderly: A randomized controlled trial. *PLoS One, 7*(1), Article e29676.
Öhman, F., Hassenstab, J., Berron, D., Schöll, M., & Papp, K. V. (2021). Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alzheimer's and Dementia: The Journal of the Alzheimer's Association, 13*(1), Article e12217.
O'Keefe, J., & Nadel, L. (1979). The hippocampus as a cognitive map. *Behavioral and Brain Sciences, 2*(4), 487–494.
Pedroli, E., Greci, L., Colombo, D., Serino, S., Cipresso, P., Arlati, S., Mondellini, M., Boilini, L., Giussani, V., Goulene, K., Agostoni, M., Sacco, M., Stramba-Badiale, M., Riva, G., & Gaggioli, A. (2018). Characteristics, usability, and users experience of a system combining cognitive and physical Therapy in a virtual environment: Positive Bike. *Sensors, 18*(7).
Peek, S. T. M., Wouters, E. J. M., van Hoof, J., Luijkx, K. G., Boeije, H. R., & Vrijhoef, H. J. M. (2014). Factors influencing acceptance of technology for aging in place: A systematic review. *International Journal of Medical Informatics, 83*(4), 235–248.
Pennanen, C., Kivipelto, M., Tuomainen, S., Hartikainen, P., Hänninen, T., Laakso, M. P., Hallikainen, M., Vanhanen, M., Nissinen, A., Helkala, E.-L., Vainio, P., Vanninen, R., Partanen, K., & Soininen, H. (2004). Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiology of Aging, 25*(3), 303–310.
Pereira, P., Duarte, E., Rebelo, F., & Noriega, P. (2014). A review of gamification for health-related contexts. Design, user experience, and usability. *User Experience Design for Diverse Interaction Platforms and Environments*, 742–753.
Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., Ritchie, K., Rossor, M., Thal, L., & Winblad, B. (2001). Current concepts in mild cognitive impairment. *Archives of Neurology, 58*(12), 1985–1992.
Pinto, T. C. C., Machado, L., Bulgacov, T. M., Rodrigues-Júnior, A. L., Costa, M. L. G., Ximenes, R. C. C., & Sougey, E. B. (2019). Is the Montreal cognitive assessment (MoCA) screening superior to the Mini-mental state Examination (MMSE) in the detection of mild cognitive impairment (MCI) and Alzheimer's disease (AD) in the elderly? *International Psychogeriatrics/IPA, 31*(4), 491–504.
Plácido, J., de Almeida, C. A. B., Ferreira, J. V., de Oliveira Silva, F., Monteiro-Junior, R. S., Tangen, G. G., Laks, J., & Deslandes, A. C. (2022). Spatial navigation in older adults with mild cognitive impairment and dementia: A systematic review and meta-analysis. *Experimental Gerontology, 165*, Article 111852.
Puthusseryppady, V., Morrissey, S., Spiers, H., Patel, M., & Hornberger, M. (2022). Predicting real world spatial disorientation in Alzheimer's disease patients using virtual reality navigation tests. *Scientific Reports, 12*(1), Article 13397.
Riva, G., Baños, R. M., Botella, C., Wiederhold, B. K., & Gaggioli, A. (2012). Positive technology: Using interactive technologies to promote positive functioning. *Cyberpsychology, Behavior, and Social Networking, 15*(2), 69–77.
Robert, P. H., König, A., Amieva, H., Andrieu, S., Bremond, F., Bullock, R., Ceccaldi, M., Dubois, B., Gauthier, S., Kenigsberg, P.-A., Nave, S., Orgogozo, J. M., Piano, J., Benoit, M., Touchon, J., Vellas, B., Yesavage, J., & Manera, V. (2014). Recommendations for the use of Serious Games in people with Alzheimer's Disease, related disorders and frailty. *Frontiers in Aging Neuroscience, 6*, 54.
Rogers, W. A., & Fisk, A. D. (2010). Toward a psychological science of advanced technology design for older adults. The Journals of Gerontology. *Series B, Psychological Sciences and Social Sciences, 65*(6), 645–653.
Rutkowski, S., Adamczyk, M., Pastuła, A., Gos, E., Luque-Moreno, C., & Rutkowska, A. (2021). Training using a commercial immersive virtual reality system on hand–eye coordination and Reaction time in young Musicians: A pilot study. *International Journal of Environmental Research and Public Health, 18*(3), 1297.
Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational Pull of video Games: A self-Determination theory approach. *Motivation and Emotion, 30*(4), 344–360.
Sardi, L., Idri, A., & Fernández-Alemán, J. L. (2017). A systematic review of gamification in e-Health. *Journal of Biomedical Informatics, 71*, 31–48.
Schankin, A., Budde, M., Riedel, T., & Beigl, M. (2022). Psychometric Properties of the user experience questionnaire (UEQ). In *CHI Conference on Human factors in computing systems*.
Schinazi, V. R., Nardi, D., Newcombe, N. S., Shipley, T. F., & Epstein, R. A. (2013). Hippocampal size predicts rapid learning of a cognitive map in humans. *Hippocampus, 23*(6), 515–528.
Schinazi, V. R., & Thrash, T. (2018). Cognitive neuroscience of spatial and geographic thinking. In *Handbook of Behavioral and cognitive Geography*. Edward Elgar Publishing.
Schöberl, F., Pradhan, C., Irving, S., Buerger, K., Xiong, G., Kugler, G., Kohlbecher, S., Engmann, J., Werner, P., Brendel, M., Schneider, E., Perneczky, R., Jahn, K., Fougère, C. la, Bartenstein, P., Brandt, T., Dieterich, M., & Zwergal, A. (2020). Real-

space navigation testing differentiates between amyloid-positive and -negative aMCI. *Neurology, 94*(8), e861–e873.

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Construction of a benchmark for the user experience questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence, 4*(4), 40.

Seçer, İ., & Us, E.Ö. (2023). Digital gaming trends of middle-aged and older adults: A sample from Turkey. *Simulation & Gaming, 54*(1), 85–103.

Segen, V., Ying, J., Morgan, E., Brandon, M., & Wolbers, T. (2022). Path integration in normal aging and Alzheimer's disease. *Trends in Cognitive Sciences, 26*(2), 142–158.

Singh-Manoux, A., Kivimaki, M., Glymour, M. M., Elbaz, A., Berr, C., Ebmeier, K. P., Ferrie, J. E., & Dugravot, A. (2012). Timing of onset of cognitive decline: Results from Whitehall II prospective cohort study. *BMJ, 344*, Article d7622.

Soldan, A., Pettigrew, C., Zhu, Y., Wang, M.-C., Moghekar, A., Gottesman, R. F., Singh, B., Martinez, O., Fletcher, E., DeCarli, C., Albert, M., & BIOCARD Research Team. (2020). White matter hyperintensities and CSF Alzheimer disease biomarkers in preclinical Alzheimer disease. *Neurology, 94*(9), e950–e960.

Spence, I., & Feng, J. (2010). Video Games and spatial cognition. *Review of General Psychology: Journal of Division 1, of the American Psychological Association, 14*(2), 92–104.

Stangl, M., Kanitscheider, I., Riemer, M., Fiete, I., & Wolbers, T. (2020). Sources of path integration error in young and aging humans. *Nature Communications, 11*(1), 2626.

Tobler, W. R. (1965). Computation of the correspondence of geographical patterns. *Papers in Regional Science: The Journal of the Regional Science Association International, 15*(1), 131–139.

Todorov, V., & Filzmoser, P. (2009). An object-Oriented framework for robust Multivariate analysis. *Journal of Statistical Software, 32*(3), 1–47. https://www.jstatsoft.org/article/view/v032i03/.

Tsai, J.-C., Chen, C.-W., Chu, H., Yang, H.-L., Chung, M.-H., Liao, Y.-M., & Chou, K.-R. (2016). Comparing the sensitivity, specificity, and predictive values of the Montreal cognitive assessment and Mini-mental state Examination when screening people for mild cognitive impairment and dementia in Chinese population. *Archives of Psychiatric Nursing, 30*(4), 486–491.

Tu, S., Wong, S., Hodges, J. R., Irish, M., Piguet, O., & Hornberger, M. (2015). Lost in spatial translation - a novel tool to objectively assess spatial disorientation in Alzheimer's disease and frontotemporal dementia. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 67*, 83–94.

Tziraki, C., Berenbaum, R., Gross, D., Abikhzer, J., & Ben-David, B. M. (2017). Designing serious computer Games for people with moderate and advanced dementia: Interdisciplinary Theory-Driven pilot study. *JMIR Serious Games, 5*(3), Article e16.

van der Ham, I. J. M., Claessen, M. H. G., Evers, A. W. M., & van der Kuil, M. N. A. (2020). Large-scale assessment of human navigation ability across the lifespan. *Scientific Reports, 10*(1), 3299.

Velandia, P. P., Miller-Petrie, M. K., Chen, C., Chakrabarti, S., Chapin, A., Hay, S., Tsakalos, G., Wimo, A., & Dieleman, J. L. (2022). Global and regional spending on dementia care from 2000-2019 and expected future health spending scenarios from 2020-2050: An economic modelling exercise. *EClinicalMedicine, 45*, Article 101337.

Vyshedskiy, A., Netson, R., Fridberg, E., Jagadeesan, P., Arnold, M., Barnett, S., Gondalia, A., Maslova, V., de Torres, L., Ostrovsky, S., Durakovic, D., Savchenko, A., McNett, S., Kogan, M., Piryatinsky, I., & Gold, D. (2022). Boston cognitive assessment (BOCA) — a comprehensive self-administered smartphone- and computer-based at-home test for longitudinal tracking of cognitive performance. *BMC Neurology, 22*(1), 1–12.

Wallace, S. E., Donoso Brown, E. V., Simpson, R. C., D'Acunto, K., Kranjec, A., Rodgers, M., & Agostino, C. (2019). A comparison of electronic and paper versions of the Montreal cognitive assessment. *Alzheimer Disease and Associated Disorders, 33*(3), 272.

Weniger, G., Ruhleder, M., Lange, C., Wolf, S., & Irle, E. (2011). Egocentric and allocentric memory as assessed by virtual reality in individuals with amnestic mild cognitive impairment. *Neuropsychologia, 49*(3), 518–527.

Wiemeyer, J., & Kliem, A. (2011). Serious games in prevention and rehabilitation—a new panacea for elderly people? *European Review of Aging and Physical Activity: Official Journal of the European Group for Research into Elderly and Physical Activity, 9*(1), 41–50.

Wiener, J. M., Carroll, D., Moeller, S., Bibi, I., Ivanova, D., Allen, P., & Wolbers, T. (2020). A novel virtual-reality-based route-learning test suite: Assessing the effects of cognitive aging on navigation. *Behavior Research Methods, 52*(2), 630–640.

Wilson, S. A., Byrne, P., Rodgers, S. E., & Maden, M. (2022). A systematic review of smartphone and tablet Use by older adults with and without cognitive impairment. *Innovation in Aging, 6*(2).

Witmer, B. G., Jerome, C. J., & Singer, M. J. (2005). The factor Structure of the presence questionnaire. *Presence: Teleoperators and Virtual Environments, 14*(3), 298–312.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*(3), 225–240.

Wong, W. (2020). Economic burden of Alzheimer disease and managed care considerations. *American Journal of Managed Care, 26*(8 Suppl), S177–S183.

World Health Organization. (2023). In *Dementia*. https://www.who.int/news-room/fact-sheets/detail/dementia.

Xiao, Y.-M., Wang, Z.-M., Wang, M.-Z., & Lan, Y.-J. (2005). [The appraisal of reliability and validity of subjective workload assessment technique and NASA-task load index]. *Zhonghua lao dong wei sheng zhi ye bing za zhi = Zhonghua laodong weisheng zhiyebing zazhi = Chinese journal of industrial hygiene and occupational diseases, 23*(3), 178–181.

Zakzanis, K. K., Quintin, G., Graham, S. J., & Mraz, R. (2009). Age and dementia related differences in spatial navigation within an immersive virtual environment. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research, 15*(4), CR140–CR150.

Zhang, J.-X., Wang, L., Hou, H.-Y., Yue, C.-L., Wang, L., & Li, H.-J. (2021). Age-related impairment of navigation and strategy in virtual star maze. *BMC Geriatrics, 21*(1), 108.

Zygouris, S., Segkouli, S., Triantafyllidis, A., Giakoumis, D., Tsolaki, M., Votis, K., & Tzovaras, D. (2022). Usability of the virtual supermarket test for older adults with and without cognitive impairment. *Journal of Alzheimer's Disease Reports, 6*(1), 229–234.