Int. J. Human–Computer Studies 167 (2022) 102885

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

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

The cupboard task: An immersive virtual reality-based system for everyday memory assessment

José Varela-Aldás [a,b], Jorge Buele [a,b], Rebecca Amariglio [c,d], Iván García-Magariño [e], Guillermo Palacios-Navarro [a,*]

[a] Department of Electronic Engineering and Communications, University of Zaragoza, Teruel Polytechnic School of Engineering, C/ Atarazana, 2, Teruel 44003, Spain
[b] SISAu Research Group, Facultad de Ingeniería y Tecnologías de la Información y la Comunicación, Universidad Tecnológica Indoamérica, Ambato 180103, Ecuador
[c] Medical School, Harvard University, Boston, MA 02115, USA
[d] Massachusetts General Hospital, Boston, MA 02114, USA
[e] Department of Software Engineering and Artificial Intelligence, Complutense University of Madrid, Madrid 28040, Spain

A R T I C L E I N F O

Keywords:
Cognitive rehabilitation
Head-mounted display
Immersive virtual reality
Memory assessment
Loci method

A B S T R A C T

Background and objective: Virtual Reality (VR) has the capacity to be used in cognitive rehabilitation interventions for diagnostic and training purposes. This technology allows the development of proposals that traditionally have been only implemented using physical elements that imply greater resources and a lesser degree of automation. This work presents an immersive virtual reality (IVR) application (the Cupboard task) for the evaluation of memory in a more ecological way and based on an activity of daily living (ADL).
Methods: To appraise its construct validity, we have carried out a comparative study with a traditional method of memory assessment (method of loci). To check for any association between performance and age, performance with years of education, and reaction time with age, the Pearson's correlation was used. One-way ANOVA was used to check for differences in performance by gender. We also performed a reliability analysis with a two way mixed effects model where people effects are random and measures effects are fixed. Therefore, intra-class correlation coefficient with absolute agreement was reckoned to assess the consistency or concordance of the measures made by both the method of loci and the cupboard IVR task.
Results: Both tasks were evaluated on a sample of 22 healthy participants who voluntarily took part in the experiment. The results obtained showed a high degree of concordance between both memory performance measures, which assumes good clinical relevance. In addition, other age-related effects were found, common to memory assessment tasks.
Conclusions: This work showed that it is possible to use an IVR application to successfully assess everyday memory. We have also demonstrated the potential of IVR to develop valid tests that assess memory functions reliably and efficiently and within ecologically valid contexts. The results obtained open the door to its use in clinical settings for cognitive training (and promoting cognitive health) of patients with mild cognitive impairment (MCI), severe cognitive impairment (SCI) such as Alzheimer or Dementia, etc., with full guarantees of application, although it must first be validated through a randomized control trial (RCT). The degree of usability of the Cupboard task was very high according to the test carried out by the participants.

1. Introduction

The use of VR is widely accepted in clinical settings that allow neuropsychology to treat various cognitive problems (Varela-Aldás et al., 2021). It can also be used to treat cognitive impairment caused by Traumatic brain injury (TBI) (Alashram et al., 2019). Likewise, there are other age-related diseases, such as MCI, which have also been treated through VR. Some of the positive effects that this technology has introduced have been shown in (Gamito et al., 2019; Liao et al., 2020). In the case of major cognitive impairment, like dementia, VR environments have been also developed (D'Cunha et al., 2019; Thapa et al., 2020).

Tests, tools and procedures that have been traditionally performed in a manual way can now be automated with the use of new technologies.

* Corresponding author.
E-mail address: guillermo.palacios@unizar.es (G. Palacios-Navarro).

https://doi.org/10.1016/j.ijhcs.2022.102885
Received 4 September 2021; Received in revised form 21 June 2022; Accepted 30 June 2022
Available online 14 July 2022

VR represents an affordable approach to recovering cognitive functions (Fernandez Montenegro and Argyriou, 2017). There are also other methods for memory training using VR, such as the method of loci (or memory palace method). It is a memorization technique whose origin dates back to classical Greece and Rome and which basically consists of assigning the different parts of the text that you want to memorize to different physical places (real or imaginary). Several studies (Peeters and Segundo-Ortin, 2019; Reggente et al., 2020; Krokos, Plaisant, and Varshney, 2019) have presented different immersive proposals based on this technique that seek to give it greater dynamism and therefore better results.

Another aspect of human cognition that can be affected deals with the visuospatial functions, which allow the representation, analysis and manipulation of objects mentally. In particular, although these functions allow the individual to interact with three-dimensional objects, a great deal of research has been carried out only in two dimensions (Korečko et al., 2019). Thus, it is relevant to determine whether the degree of immersion and the differences between the two and three dimensional designs have an influence on the rehabilitation processes. Ventura et al. (2019) evaluated memory performance through two exercises, the first one within an immersive environment and the second one within a non-immersive environment. The results showed that the group that began with the immersive exercise obtained a better performance when evaluating cognitive function, compared to the other group.

We find different studies in literature that have used ecological EVs for cognitive training and memory evaluation to detect memory dysfunctions (Ruet and Brochet, 2020). For example, Ouellet et al. (2018) focused on the evaluation of everyday memory, whereas Pflueger et al. (2018) and Moffat et al. (2001) evaluated spatial memory. Jebara's study did the same with episodic memory (Jebara et al., 2014) and attention and executive functioning were addressed in the study of Armstrong et al. (2013). We also found examples that have focused on patients with multiple sclerosis with cognitive deficiencies (Goverover et al., 2010; Lamargue-Hamel et al., 2015) or executive disorders (Rouaud et al., 2006). Most of the studies have concluded that these tools are perfectly valid not only for cognitive training in older adults (Corriveau Lecavalier et al., 2020), but for improving memory functions in brain-damaged adults (Caglio et al., 2012) or even for the assessment of episodic memory in Alzheimer's patients (Lecouvey et al., 2019), among others. Furthermore, they have the advantages (with respect to traditional test batteries) of providing an objective measure of individual memory components in ADL simulations (Parsons and McMahan, 2017). To see the importance of the design of this tools compared to the classical evaluation through batteries of neuropsychological tests, some authors have demonstrated their ability to detect executive dysfunctions in daily live conditions that were underestimated by these classical tests in a group of patients with multiple sclerosis (Rouaud et al., 2006). This can be partly understandable since the classic test batteries are performed under very simple and artificial conditions with few or no distractors (Ruet and Brochet, 2020).

In this context, some of the abovementioned studies have included tasks that are based on activities that users can perform in their daily life. The activities of daily living (ADLs) are often called basic ADLs, and they include the fundamental skills typically needed to manage basic physical need, belonging to areas such as personal hygiene, dressing, toileting, ambulating, and eating (Mlinac and Feng, 2016). On the other hand, we can make the distinction with the instrumental ADLs (IADLs), which include activities of a higher level and related to independent living in community. According to Lawton and Brody (1969), these IADLs include activities such as ability to use the telephone, shopping, food preparation, housekeeping, laundry, mode of transportation, responsibility for own medications and the ability to handle finances. However, the concept of IADL should be extended to social participation, since it is closely linked to it. In fact, in the study of Tomioka et al. (2016), a strong association was found between participation in social groups and performing IADLs independently among the community-dwelling elderly, regardless of gender.

The analysis of the literature has shown us the usefulness of VR systems in cognitive training and assessment processes. However, the analyzed works have been more focused on usability and reliability (Krokos et al., 2019; Ventura et al., 2019,; Jebara et al., 2014; Corriveau Lecavalier et al., 2020) without validating the performance with traditional and clinically validated tools. Other works have focused on the evaluation of the performance of everyday life activities (Goverover et al., 2010; Lamargue-Hamel et al., 2015) not being memory tasks per se, for example, purchasing an airline ticket online using the Internet (Goverover et al., 2010; Lamargue-Hamel et al., 2015), or a driving simulator task (Goverover et al., 2010; Lamargue-Hamel et al., 2015). Some works (Ouellet et al., 2018; Moffat et al., 2001; Armstrong et al., 2013; Rouaud et al., 2006; Parsons and McMahan, 2017) have validated performance results with paper-and-pencil test measures by means of correlation coefficient, without using more reliable validation measures such as the intraclass correlation coefficient. Therefore, we consider that there is still a gap that we intend to fill with this work: the use of immersive environments to develop IADL-based memory tasks that can be validated with reliable clinical instruments.

In this work, a system for everyday memory assessment is proposed, using an immersive virtual reality (IVR) environment. Thus, the user exercises sensory-visual memory, working memory, spatial orientation and executive function at the same time. The work is based on an experiment carried out by González-Landero et al. (2019), who proposed a prototype of an automated three-door cupboard with Internet of Things (IoT) technology. The system now proposed fully virtualizes this prototype and increases the cupboard with one more door (4 in total), making data collection (response times and accuracy response) in a more automated and simple way. We hypothesized that our IVR application is an ecological task that can be effective for the assessment of everyday memory. Our hypothesis will be validated with a traditional method such as the method of loci. At the same time, we will check if the task is sensitive to the typical age-related differences effect observed on episodic memory tasks.

2. Materials and methods

2.1. Design of the virtual environment

In this work, we intend to bring to VR an instrumental ADL such as the opening of a series of cupboards to take different objects that can be found in them so that it constitutes a memory assessment tool. We have chosen an IADL task because the associated functions are normally lost before basic ADLs, and therefore the assessment of IADLs may identify early decline (Ward et al., 1998). Fig. 1 shows the general scheme of the IVR-based system. Immersion is achieved using a Head-mounted display (HMD) device (Oculus Go) and a handheld controller that allows interaction with the system. The results generated are stored in the local database and sent to a remote server on the network, for a subsequent analysis. Unity has been used as a game engine.

The virtual environment consists of a task (cupboard task) that is based on an activity of daily life (ADL) for the assessment of everyday memory. Fig. 2 illustrates the two stages of the game. In the training stage (study phase) the user visualizes the elements inside the cupboards. The elements are placed automatically and randomly within the shelves and the doors can be opened individually, with a waiting time for learning that has been set at 10 s (González-Landero et al., 2019). In the evaluation stage (recall phase), the user must locate the elements requested by the application, by selecting both the door and the desired element. The user has a time limit to choose a door and an element (10 s), otherwise an error is recorded and the evaluation continues. Accuracy responses and response times were recorded in plain text files to be saved in the local database and sent to the remote server (SFTP). The data is first stored in a local database to protect us against eventual drops in the WIFI signal in the place where the experiment is carried out. In the

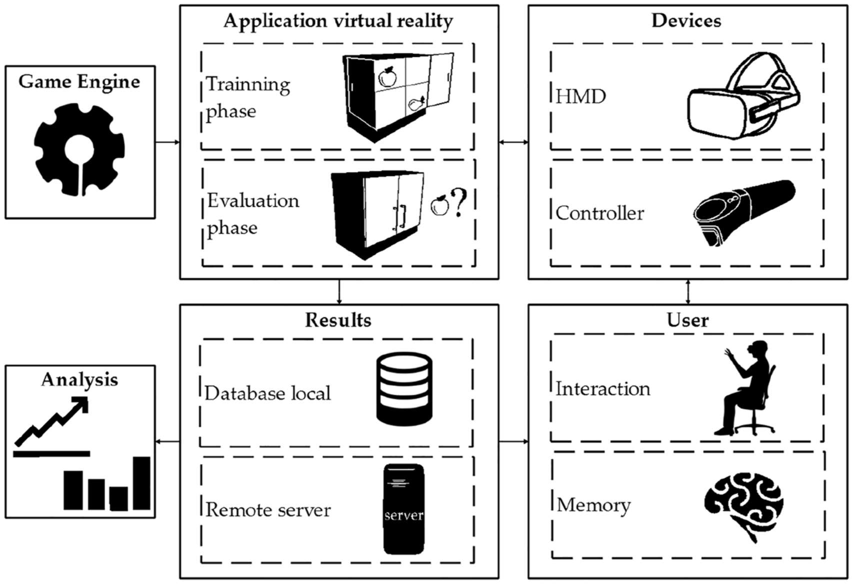

Fig. 1. General description of the IVR-based system.

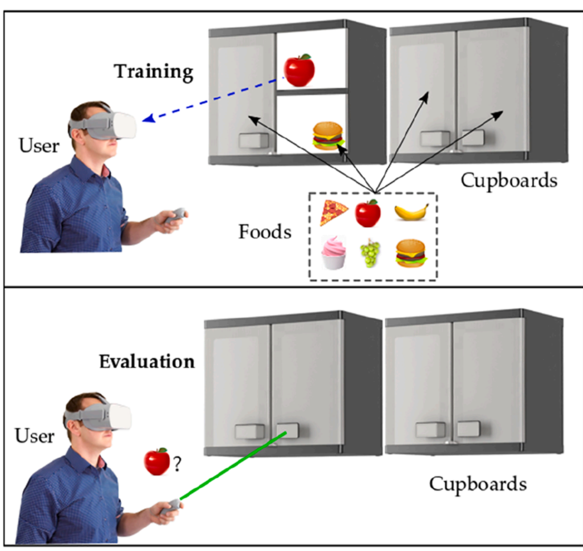

Fig. 2. Activities to be implemented in the VR application.

event of such drops or a poor signal, the data will be stored first in the local database and the next time a proper connection is provided, the data will immediately be sent to the remote server.

2.2. Tasks

To evaluate our IVR-based proposal, we implemented a task based on the method of loci. To apply the technique, each participant had to imagine a house with four rooms to act as the Memory Palace, where different objects were placed in each room. The participant was asked for the places where the elements were placed. The Memory Palace task was implemented through a non-immersive approach (by means of a desktop computer application and a mouse-based interaction). On the other hand, the Cupboard task was implemented via an immersive approach through HMD.

The participants were asked to answer the questions in the shortest time possible. Both tasks were designed according to two levels of difficulty, by configuring 7 elements for the first level and 9 elements for the second level. This means that tthe number of elements to memorize was doubled (14 and 18, respectively). The main reason for selecting such number of objects is twofold. On one hand, it has to do with the experience of other studies dealing with memory tasks. Similar studies such as the one conducted by Plechatá et al., 2021 used four different levels of difficulty with 3, 5, 7 and 9 items to memorize, respectively. Other authors decided to work with a set of 21 faces to remember (Krokos et al., 2019). Reggente et al. (2020) used a sequence of 15 3D objects whereas Corriveau Lecavalier et al. (2020) used 12 articles in his virtual store task and later recovered them in the recall phase from among 24 objects. Although it was not a virtual task, González-Landero et al. (2019) used 15 objects to memorize in a 3-compartment physical cupboard. On the other hand, since we worked in a virtual environment the objects had to be placed in the different compartments with a reasonable level of resolution according to the cupboard dimensions. Finally, the design of interventions with different levels of difficulty has previously been used in numerous cognitive training studies (Gamito et al., 2019; Aarnoudse-Moens et al., 2018; Gamito et al., 2020).

2.3. Intervention

The experiment took place during the covid-19 pandemic, where the mobility conditions for the dates of the experiments were severely restricted. This resulted in a very hard recruiting process. Although an RCT may provide more evidence, we opted for a quasi-experimental one-group design, in order to seek greater efficiency of the statistical contrast by reducing variability and so increasing statistical power.

The experiment was performed in two different days. The first day was devoted to the Loci task, whereas the second day the cupboard task was performed. Before carrying out the Loci task, each participant worked on concentration exercises for 15 min to prepare them for the task. In the VR task, users were familiarized with the virtual environment prior to the start of the tests for 15 min approximately (handling of the laser pointer to open doors, selection of objects, etc.). Therefore, the intervention time for every participant was about 30 min each day (60 min for the whole experiment). Fig. 3 shows a participant performing the Cupboard task with the Oculus Go.

2.4. Statistical analysis

Statistical analysis was carried out using SPSS 16. Shapiro-Wilk test was performed to check for data normality (the sample is less than 30 participants). Pearson's Chi-squared test was used to verify the

Fig. 3. Participant performing the Cupboard task with the Oculus Go.

homogeneity regarding to gender. To check for any association between performance and age, performance with years of education, and reaction time with age, the Pearson's correlation was used. One-way ANOVA was used to check for differences in performance by gender. We also performed a reliability analysis with a two way mixed effects model where people effects are random and measures effects are fixed. Therefore, intra-class correlation coefficient with absolute agreement was reckoned to assess the consistency or concordance of the measures made by both the Memory Palace method and the cupboard IVR task. Finally, a usability test was completed by the sample to measure feasibility of the developed IVR task. The statistical significance level was set to 0.05.

2.5. Participants

Initially we wanted to detect a large correlation according to Cohen (1988) (≥ 0.5), between both methods. Therefore, for a significance level equal to 0.05 and a power of 80% (the usual value in research), GPower software (G Power, 2020) gave us a sample of 21 subjects. A total of 22 healthy subjects were recruited. The Mini Mental State Examination (MMSE) test was used as a screening measure in the enrollment phase. The inclusion criteria were: (a) ages between 18 and 65 years old; (b) MMSE score greater or equal to 26 as suggested by Kukull et al. (1994); and (c) no physical limitations, especially in the hand they use most frequently. Initial training was given to the participants prior to the experiment to give them the opportunity to get familiar with the VR setup (understanding of controls and basic actions). The study was carried out according to the Declaration of Helsinki and was approved by the Institutional Review Board of Universidad Tecnológica Indoamerica (protocol code UTI-IIDI-054-2021, approval date 18 March 2021). All the participants were volunteers and signed an informed consent previous to their participation in the research. The demographics of the participants is depicted in Table 1. No significant differences were found in the sample with respect to gender (p = .394).

3. Results

We present the results obtained after the experiments for both tasks, as well as a comparison in terms of accuracy response. Table 2 shows the accuracy response results as a percentage for both tasks (memory palace and VR Cupboard).

Table 1
Participant demographics.

Demographics	Value	%	Demographics	Value
Gender:			Age	
Male	9	40.91	Mean	36.64
Female	13	59.09	SD	12.91
Experience VR			Education years	
Yes	7	31.82	Mean	13.59
No	15	68.18	SD	4.24
Cognitive exercise				
Yes	17	77.27		
No	5	22.73		

3.1. Memory palace

The results obtained after applying the memory palace method collected the precision and reaction times of the 22 participants in 2 levels of difficulty (with 2 rounds per level). Although a positive and high significant association (r = 0.477) was observed in the accuracy response in the two levels, the paired-samples t test did not reveal significant differences between levels (t (21), p = 0.219). No association was found between performance and age (Pearson's correlation not statistically significant) either. Regarding reaction times, we only found a significant positive correlation with age only in the case of 9 elements (r = 0.551, p = 0.008).

3.2. Cupboard task

The procedure to be carried out in the Cupboard task was similar to the previous one. With regard to the relationship between performance and age, a very high and negative Pearson correlation (- 0.751) significant at the 0.01 level was found. On the contrary, we did not find any association between performance and years of formal education (p > 0.05). There were no differences by sex in terms of performance at this level either (F (1.20) = 0.379, p = 0.545)). Regarding reaction times, three were recorded: the cupboard time (CT) or time to select the cupboard, the object time (OT) or time it takes to select the object, and the total time (TT), or sum of the previous times.

On the other hand, and for level 1, we found a high and positive association (Pearson's coefficient = 0.814) significant at the 0.01 level between the cupboard reaction time and the age of the participants. We repeated the calculation for the total reaction time of the task, and we find a very similar situation (Pearson's coefficient high (0.831) and significant at the 0.01 level). We also had a negative and high association between total reaction time and performance (Pearson's coefficient = -0.928, p < 0.001), which indicates that the longer the reaction time, the worse the results. Regarding the relationship between performance and age, no significant Pearson correlation was found. Similarly, no association was found between performance and years of formal education (p > 0.05). Therefore, the performance in the execution of the task (in the two levels of difficulty) does not depend on the educational level of the participants. We did not find differences by sex in terms of performance at this level (F (1,20) = 2.413, p = 0.136)) either. The paired-samples t test did not reveal significant differences between the accuracy response at the two levels (t (21), p = 0.795).

In relation to reaction times (cupboard time) and age, we found a moderate and positive association (Pearson's coefficient = 0.457) significant. Although in the case of 9 objects the correlation is lower than in the case of 7, in both cases it is observed that the older the subject, the longer the response time is. We found a very similar situation (Pearson's coefficient =0.504 and significant p = 0.017) after repeating the calculation for the total time. We also had a negative and high association between total reaction time and performance (Pearson's coefficient = -0.690, p < 0.001), which indicated that the longer the reaction time, the worse the results.

Table 2
Accuracy response results (%) for both tasks (Memory Palace and VR Cupboard).

	Measure	Memory Palace [%] Round 1	Round 2	All	VR Cupboard task [%] Round 1	Round 2	All
Level 1	Mean	62.5	71.99	67.24	71.38	70.86	71.12
	SD	25.61	27.81	26.85	20.91	17.47	19.04
Level 2	Mean	69.7	76.26	72.98	70.48	70.66	70.57
	SD	24.77	20.94	22.91	16.32	14.99	15.49

3.3. Accuracy response. Comparisons between methods

In this section we compare the two methods for memory assessment. To assess the consistency of the measures made by both the Memory Palace method and the cupboard task and thereby validate the latter, we have reckoned the intra-class correlation coefficient with absolute agreement. This measure is much more robust than Pearson's correlation coefficient, since the latter does not assess agreement since it is not sensitive to differences of constant type or proportional type, respectively.

For the first level (7 objects) an intra-class correlation coefficient of .614 was obtained, while for the second level (9 objects) an intra-class correlation coefficient of, 0.662 was obtained, which according to Cicchetti (1994), correspond to a good level of clinical significance (reliability coefficient between .60 and .74). In addition, in both cases the result of the ANOVA analysis reflects the absence of bias because the means obtained with these two tools do not present significant differences ($F = 0.507$; $p = 0.484$) for the case of 7 items and ($F = 0.917$; $p = 0.349$) for the case of 9 items, respectively.

To complement the validity analysis of our method, and since the differences in measurements between the two methods followed a normal distribution (Shapiro-Wilk test non-significant), we were able to apply the Bland-Atman graphical test (Bland-Atman plot), whose result is depicted in Figs. 4 and 5, respectively. We see that the mean difference is close to zero, -3.32 at the first level (see Fig. 4) and 3.2 at the second level (see Fig. 5), respectively, which indicates that both methods obtain similar means. For the first level (7 objects), the VR Cupboard method overestimates the results with respect to the Loci method (negative mean), since it measures 3.32 more units on average. On the other hand, in the second level (9 objects), the VR Cupboard method underestimates the results (positive mean), since it measures 3.2 units less (on average). Anyway, such value (systematic bias) is limited to ±3%, approximately, with respect to the maximum of the scale, which implies a good approximation of the methods. It is observed that 95% of the differences between both methods are located between the limits of agreement.

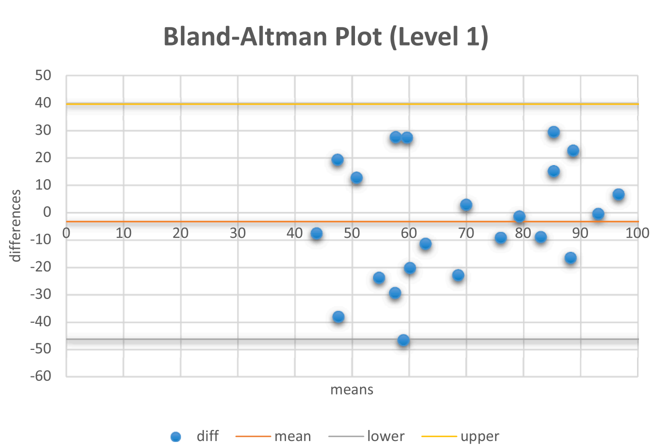

Fig. 4. Bland-Altman plot for both Memory Palace and VR Cupboard task in the first level of difficulty (7 objects). Horizontal axis: mean of both methods. Vertical axis: differences between methods (paired values). Lower and upper limits of agreement mean 95% CI for differences between methods.

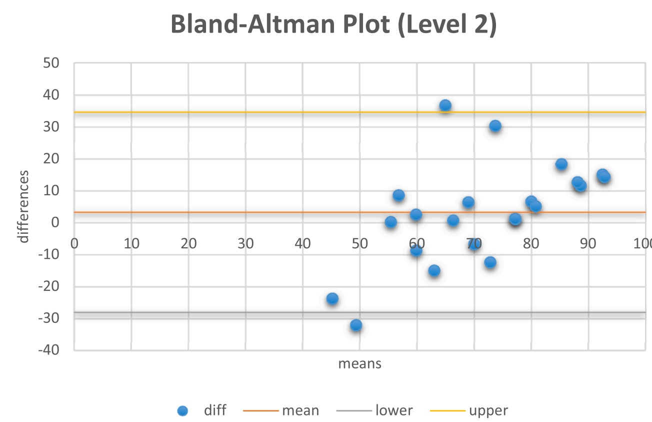

Fig. 5. Bland-Altman plot for both Memory Palace and VR Cupboard task in the second level of difficulty (9 objects). Horizontal axis: mean of both methods. Vertical axis: differences between methods (paired values). Lower and upper limits of agreement mean 95% CI for differences between methods.

Only a couple of outliers appear at level 2 (9 items), which may be due to the increased difficulty. These limits establish the range in which the differences will be found in the data of both methods (approximately 95% of the time). Finally, it is important to mention that the variability is consistent throughout the range of values and around the bias line (mean difference line).

3.4. Usability analysis

To assess the feasibility of the application from the participants' point of view, the USEQ (User Satisfaction Evaluation Questionnaire) (Gil-Gómez et al., 2017) satisfaction test was applied. The results showed us a high degree of participant satisfaction (see Table 3). In general, the participants approved the characteristics of the system and considered that the proposal may be helpful in cognitive rehabilitation. An additional question about their task preference (Memory Palace vs Cupboard Task) was asked to the participants, and 86.36% of them showed preference for the Cupboard Task.

Table 3
USEQ test results for the VR Cupboard task.

Question	Mean	SD
Q1. Did you enjoy your experience with the system?	4.82	0.49
Q2. Were you successful using the system?	4.43	0.53
Q3. Were you able to control the system?	4.91	0.29
Q4. Is the information provided by the system clear?	4.95	0.21
Q5. Did you feel discomfort during your experience with the system?	4.82	0.49
Q6. Do you think that this system will be helpful for rehabilitation?	4.95	0.21
Total:	28.89/30	0.37

4. Discussion

4.1. Construct validity

Our hypothesis regarding the construct validity of the VR-based designed task was that the memory performance in the Cupboard task would have a great deal of concordance to that obtained via a traditional assessment method, such as the method of loci. The results obtained (accuracy response) after comparison of the developed system with a conventional memory measurement system (method of memory palace or loci), attest to this. The measure of consistency or concordance (with absolute agreement) of both methods, performed through the intra-class correlation coefficient, has revealed a good level of clinical significance, according to Cicchetti (1994). This was true for the two levels of difficulty implemented in the task. Moreover, no bias was observed between the means obtained by both methods and both levels of difficulty. The choice of this measure of concordance, through ICC (instead of Pearson's correlation), together with the verification through the Bland-Atman diagram, provide us with a high degree of validity in the confirmation of our hypothesis. Although our results align with previous studies pointing out that VR can measure similar constructs as those measured by clinical or experimental measures, we find some differences with them.

Taking into account similar studies identified in the literature, some of them do not make any comparison between the performance of both methods (VR task versus traditional cognitive tests) to validate the experimental method (Goverover et al., 2010; Lamargue-Hamel et al., 2015; Plechatá et al., 2021). Other studies use the Pearson's correlation to see the degree of association between the cognitive performance achieved with the virtual tool and some traditional cognitive tests (Ouellet et al., 2018; Moffat et al., 2001; Parsons and McMahan, 2017). For example, Moffat et al. (2001) found moderate to low correlations between navigation performance in the virtual task and traditional psychometric measures of cognitive aging. Parsons and McMahan (2017) only found moderate correlations with the traditional neuropsychological memory scores in her Virtual Environment Grocery Store task for both conditions imposed in their study (low distraction and high distraction on the virtual task). Ouellet et al. (2018) only found moderate correlations in the scores for correct recall in the Virtual Shop task compared to traditional neuropsychological measure of episodic memory ($r=0.35$). However, they found no association between performance on the measure of episodic memory and the Multifactorial Memory Questionnaire (Troyer et al., 2006). Therefore, the degrees of association found by the aforementioned studies were moderate to low, whereas our study found a good clinical validity in our comparison through an ICC greater than 0.6.

We also think that the good results obtained by our task could have come from the fact that the task, as occurs with the loci method, reinforces an object-location binding behavior during encoding (Reggente et al., 2020). This is one of the strengths that we have tried to exploit in the VR Cupboard task, which reinforces theoretical memory models that emphasize spatial context in encoding as the core of mnemonic function (Reggente et al., 2020) and adds evidence to the fact that spatial context has influence on event memory (Robin et al., 2018). Some authors such as Wallet et al. (2011) have suggested that spatial knowledge from virtual environments could be transferable to the real world.

Another aspect that has been widely discussed in the literature and that has been able to contribute to obtaining good results in our study lies in the fact that the assessment task is based on an instrumental activity of daily living (IADL). Our results may have been influenced (positively) by the inclusion of a more meaningful environment as evidence suggests that memory for real-life objects is more efficient when the objects are in a context (Pflueger et al., 2018). This makes this type of memory assessment tasks closer to the functioning of everyday memory and therefore can better predict problems related to memory than clinical tests, as pointed out by Pflueger et al. (2018). Also the elicitation of a strong sense of "being physically present", as well as the induction in the observer of virtual embodiment that these immersive systems provide, has also been able to contribute to the good results of our task.

4.2. Age-related differences

In addition to the above, we wanted to demonstrate that the task would be sensitive to typical age-related differences effect observed on episodic memory tasks. In fact, our task was sensitive to age, as younger adults performed better than older adults as far as the accuracy response was concerned. This means that the VR task thus designed is sensitive to typical memory impairment (Craik, 1986). These results align with other recent and related studies on VR tasks for memory assessment (Corriveau Lecavalier et al., 2020; Pflueger et al., 2018). The study of Plechatá et al. (2021) implemented the same task (shopping task) in two systems (immersive and non-immersive) and found age-related differences in terms of both assessed performance and user experience, respectively, when dealing with episodic memory. Other studies have corroborated this same age-related decrease in performance when dealing with spatial memory (Moffat, Zonderman, and Resnick, 2001; Maidenbaum et al., 2019; Coutrot et al., 2018). Regarding reaction times, the observed correlation between age and reaction times on our task aligns with other visual working memory tasks (García-Magariño et al., 2020). In the work of Armstrong et al. (2013) there was also a positive correlation of age with the average response times for the two VR conditions in their experiment. Ouellet et al. (2018) also found longer completion times in older adults compared to younger ones.

4.3. Influence of the type of virtual environment

Another noteworthy issue that we have not addressed because it was not the subject of study, deals with the influence of the type of virtual environment used (immersive vs. non-immersive) on performance in memory tasks. Krokos et al. (2019) found that the participants in the immersive condition performed 8.8% higher compared to the non-immersive condition as far as the overall average recall performance was concerned. Similarly, Huttner et al. (2019) concluded that all the average scores in the immersive condition were approximately 5 to 7% higher than those in the non-immersive condition. However, Maidenbaum et al. (2019) found that humans display similar spatial memory performance in a spatial memory task both when performing it via a standard computer screen and via an HMD. In our case, we have not found significant differences in performance between the two methods (loci vs. VR Cupboard task) and conditions (immersive vs. non-immersive). We realize that the studies carried out to date are not conclusive in this matter and therefore the debate is still open and deserves due attention.

4.4. Usability

Finally, the usability analysis of the proposal yields satisfactory results meeting the expectations of the researchers (USEQ = 96.3%), because the application was designed with practical and easy-to-use features, including clear and reduced information, and limiting movements within the virtual environment to avoid discomfort in the user. These results are consistent with other proposals for immersive VR via HMD, such as that of Hassandra et al. (2021) focused on older people with MCI symptoms. Plechatá et al. (2021) also reported a very good user experience with the immersive tool, although they used an ad hoc developed usability questionnaire.

4.5. Limitations

The study carried out has numerous limitations that should be taken into account when interpreting the results. First, the experiment took place during the covid-19 pandemic, which resulted in a very hard recruiting process and it took a lot to get the final number of participants. Since all the participants were volunteers, this detail could introduce some selection bias. On the other hand, the dose used in each task might have not been sufficient to achieve a greater improvement in performance of participants. In addition, each task was performed only once by each participant. Participants could have repeated the task after an interval of 3 weeks, thereby making some assessment of reliability through test-retest analysis. Another limitation that could be addressed in further studies is that the recall process was tested immediately after the encoding period, so we do not have a measure of long-term memory durability. As far as technical aspects of the system are concerned, some of the participants experienced some difficulties in handling the small joystick that comes with the VR googles. Hence, the reaction times have been longer. To overcome this limitation, we are working in a new version of the system by using the Oculus Quest 2, a new generation of HMDs. that allow the user interact with the VR environment with his/her hands, yet they provide full hand tracking. On one hand, the participants would not be worried and focused on pressing the right button anymore, and at the same time kinematics parameters such us reaction times will improve as they would be lower.

4.6. Applications and future work

The low cost and simplicity of our solution, together with the automation of data collection immediately, open the doors to its use as a routine memory task for the early detection of cognitive decline, not only in home-based environments, but in clinical settings. In a very simple way, and with the designed task included in a rehabilitation program, a routine that sends data to the occupational therapist to know the cognitive status of each patient can be easily programmed. By contrast, we have found in literature virtual tasks that are not easy to perform in clinical practice, since they require specific facilities such as sleep laboratories facilities. We found an example of such tasks in the one designed by Lamargue-Hamel et al. (2015) (Driving Simulator Dual Task), in which, in addition, some participants experienced motion sickness and were not able to complete the task.

Future work deals with the design of proposals for both cognitive assessment and training using immersive VR systems, to elucidate the ability of these systems to improve the cognitive performance of population. We also want to investigate of the effects of active navigation in virtual environments on memory performance. The term active navigation (mainly in virtual environments) refers to the idea of navigating through an environment in an active, self-directed way, or by performing a free exploration. Research in this area suggests that VE exploration by means of an active navigation enables superior spatial learning compared with a more passive, or observational encounter of the same environment (von Stülpnagel and Steffens, 2012). According to Wilson et al. (1997), active navigation in VE can be divided into physical activity (motor control) and psychological activity (decision-making). The processes involved in active navigation constitutes a type of enactment that can lead to memory performance enhances both in spatial (Brooks, 1999) and episodic memory (Jebara et al., 2014). Finally, we plan to carry on a randomized control trial (RCT) including participants with some cognitive impairments such as people suffering from Alzheimer's disease. Such study could show us more evidence about the discriminant ability of our task in the cognitive aspect.

5. Conclusions

This work showed that it is possible to use an IVR application to successfully assess everyday memory. The results of our study have demonstrated a good level of clinical significance, in addition to the fact that the task maintains the characteristics of traditional memory tasks (e.g. age-related differences). We have also demonstrated the potential of IVR to establish a new methodology to develop valid tests that assess memory functions reliably and efficiently and within ecologically valid contexts. The results obtained open the door to its use in clinical settings for cognitive training (and promoting cognitive health) of patients with MDI, DCI, Alzheimer, etc. with full guarantees of application, although it must first be validated through a randomized control trial (RCT).

Institutional review board statement

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of UNIVERSIDAD TECNOLÓGICA INDOAMÉRICA (protocol code UTI-IIDI-054-2021, approval date 18 March 2021).

Informed consent statement

Informed consent was obtained from all subjects involved in the study.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

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

Acknowledgments

The authors wish to thank the participants and the Universidad Tecnológica Indoámerica for the support provided in this research.

References

Aarnoudse-Moens, C.S.H., Twilhaar, E.S., Oosterlaan, J., van Veen, H.G., Prins, P.J.M., van Kaam, A.H.L.C., van Wassenaer-Leemhuis, A.G., 2018. Executive function computerized training in very preterm-born children: a pilot study. Games Health J. 7, 175–181. https://doi.org/10.1089/g4h.2017.0038.

Alashram, A.R., Annino, G., Padua, E., Romagnoli, C., Mercuri, N.B., 2019. Cognitive rehabilitation post traumatic brain injury: a systematic review for emerging use of virtual reality technology. J. Clin. Neurosci. 66, 209–219. https://doi.org/10.1016/j.jocn.2019.04.026.

Armstrong, C.M., Reger, G.M., Edwards, J., Rizzo, A.A., Courtney, C.G., Parsons, T.D., 2013. Validity of the Virtual Reality Stroop Task (VRST) in active duty military. J. Clin. Exp. Neuropsychol. 35, 113–123. https://doi.org/10.1080/13803395.2012.740002.

Brooks, B.M., 1999. The specificity of memory enhancement during interaction with a virtual environment. Memory 7, 65–78. https://doi.org/10.1080/741943713.

Caglio, M., Latini-Corazzini, L., D'Agata, F., Cauda, F., Sacco, K., Monteverdi, S., Zettin, M., Duca, S., Geminiani, G., 2012. Virtual navigation for memory rehabilitation in a traumatic brain injured patient. Neurocase 18, 123–131. https://doi.org/10.1080/13554794.2011.568499.

Cicchetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assess. 6, 284–290.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. L. Erlbaum Associates, Hillsdale, N.J.

Corriveau Lecavalier, N., Ouellet, É., Boller, B., Belleville, S., 2020. Use of immersive virtual reality to assess episodic memory: a validation study in older adults. Neuropsychol. Rehabil. 30, 462–480. https://doi.org/10.1080/09602011.2018.1477684.

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. Curr. Biol. 28, 2861–2866.e4. https://doi.org/10.1016/j.cub.2018.06.009.

Craik, F.I.M., 1986, "A functional account of age differences in memory," In: F. Klix and H. Hagendorf, Eds., Human Memory and Cognitive Capabilities: Mechanisms and Performances, Elsevier, Amsterdam, pp. 409-422.

D'Cunha, N.M., Nguyen, D., Naumovski, N., McKune, A.J., Kellett, J., Georgousopoulou, E.N., Frost, J., Isbel, S., 2019. A mini-review of virtual reality-based interventions to promote well-being for people living with dementia and mild cognitive impairment. Gerontology 65, 430–440. https://doi.org/10.1159/000500040.

Fernandez Montenegro, J.M., Argyriou, V., 2017. Cognitive evaluation for the diagnosis of Alzheimer's disease based on Turing Test and Virtual Environments. Physiol. Behav. 173, 42–51. https://doi.org/10.1016/j.physbeh.2017.01.034.

G Power, n.d. 2020 https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.

Gamito, P., Oliveira, J., Alves, C., Santos, N., Coelho, C., Brito, R., 2020. Virtual reality-based cognitive stimulation to improve cognitive functioning in community elderly: a controlled study. Cyberpsychol. Behav. Soc. Netw. 23, 150–156. https://doi.org/10.1089/cyber.2019.0271.

Gamito, P., Oliveira, J., Morais, D., Coelho, C., Santos, N., Alves, C., Galamba, A., Soeiro, M., Yerra, M., French, H., Talmers, L., Gomes, T., Brito, R., 2019. Cognitive stimulation of elderly individuals with instrumental virtual reality-based activities of daily life: pre-post treatment study. Cyberpsychol. Behav. Soc. Netw. https://doi.org/10.1089/cyber.2017.0679.

García-Magariño, I., Fox-Fuller, J.T., Palacios-Navarro, G., Baena, A., Quiroz, Y.T., 2020. Visual working memory for semantically related objects in healthy adults. Rev. Neurol 71, 277–284.

Gil-Gómez, J.A., Manzano-Hernández, P., Albiol-Pérez, S., Aula-Valero, C., Gil-Gómez, H., Lozano-Quilis, J.A., 2017. USEQ: a short questionnaire for satisfaction evaluation of virtual rehabilitation systems. Sensors 17, 1–12. https://doi.org/10.3390/s17071589.

González-Landero, F., García-Magariño, I., Amariglio, R., Lacuesta, R., 2019. Smart cupboard for assessing memory in home environment. Sensors 19, 2552. https://doi.org/10.3390/s19112552.

González-Landero, F., García-Magariño, I., Amariglio, R., Lacuesta, R., 2019. Smart cupboard for assessing memory in home environment. Sensors 19, 1–21. https://doi.org/10.3390/s19112552.

Goverover, Y., O'Brien, A.R., Moore, N.B., DeLuca, J., 2010. Actual reality: a new approach to functional assessment in persons with multiple sclerosis. Arch. Phys. Med. Rehabil. 91, 252–260. https://doi.org/10.1016/j.apmr.2009.09.022.

Hassandra, M., Galanis, E., Hatzigeorgiadis, A., Goudas, M., Mouzakidis, C., Karathanasi, E.M., Petridou, N., Tsolaki, M., Zikas, P., Evangelou, G., Papagiannakis, G., Bellis, G., Kokkotis, C., Panagiotopoulos, S.R., Giakas, G., Theodorakis, Y., 2021. A virtual reality app for physical and cognitive training of older people with mild cognitive impairment: mixed methods feasibility study. JMIR Serious Games 9, e24170. https://doi.org/10.2196/24170.

Huttner, J.-P., Qian, Z., Robra-Bissantz, S., 2019. A Virtual memory palace and the user's awareness of the method of Loci. In: Proceedings of the European Conference on Information Systems (ECIS2019. Stockholm-Uppsala, Sweden, pp. 1–14.

Jebara, N., Orriols, E., Zaoui, M., Berthoz, A., Piolino, P., 2014. Effects of enactment in episodic memory: a pilot virtual reality study with young and elderly adults. Front. Aging Neurosci 6. https://doi.org/10.3389/fnagi.2014.00338.

Korečko, Š., Hudák, M., Sobota, B., Marko, M., Cimrová, B., Farkaš, I., Rosipal, R., 2019. Assessment and training of visuospatial cognitive functions in virtual reality: proposal and perspective. In: Proceedings of the 9th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2018. Institute of Electrical and Electronics Engineers Inc., pp. 39–44

Krokos, E., Plaisant, C., Varshney, A., 2019. Virtual memory palaces: immersion aids recall. Virtual Real. 23, 1–15. https://doi.org/10.1007/s10055-018-0346-3.

Kukull, W.A., Larson, E.B., Teri, L., Bowen, J., McCormick, W., Pfanschmidt, M.L., 1994. The mini-mental state examination score and the clinical diagnosis of dementia. J. Clin. Epidemiol. 47, 1061–1067. https://doi.org/10.1016/0895-4356(94)90122-8.

Lamargue-Hamel, D., Deloire, M., Saubusse, A., Ruet, A., Taillard, J., Philip, P., Brochet, B., 2015. Cognitive evaluation by tasks in a virtual reality environment in multiple sclerosis. J. Neurol. Sci. 359, 94–99. https://doi.org/10.1016/j.jns.2015.10.039.

Lawton, M.P., Brody, E.M., 1969. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 9, 179–186.

Lecouvey, G., Morand, A., Gonneaud, J., Piolino, P., Orriols, E., Pélerin, A., Da Silva, L.F., de La Sayette, V., Eustache, F., Desgranges, B., 2019. An impairment of prospective memory in mild Alzheimer's disease: a ride in a virtual town. Front. Psychol. 10, 1–12. https://doi.org/10.3389/fpsyg.2019.00241.

Liao, Y.Y., Tseng, H.Y., Lin, Y.J., Wang, C.J., Hsu, W.C., 2020. Using virtual reality-based training to improve cognitive function, instrumental activities of daily living and neural efficiency in older adults with mild cognitive impairment. Eur. J. Phys. Rehabil. Med. 56, 47–57. https://doi.org/10.23736/S1973-9087.19.05899-4.

Maidenbaum, S., Patel, A., Stein, E., Jacobs, J., 2019. Spatial memory rehabilitation in virtual reality - extending findings from epilepsy patients to the general population. In: Proceedings of the International Conference on Virtual Rehabilitation. ICVR, Institute of Electrical and Electronics Engineers Inc.

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

Moffat, S.D., Zonderman, A.B., Resnick, S.M., 2001. Age differences in spatial memory in a virtual environment navigation task. Neurobiol. Aging 22, 787–796. https://doi.org/10.1016/S0197-4580(01)00251-2.

Ouellet, É., Boller, B., Corriveau-Lecavalier, N., Cloutier, S., Belleville, S., 2018. The virtual shop: a new immersive virtual reality environment and scenario for the assessment of everyday memory. J. Neurosci. Methods 303, 126–135. https://doi.org/10.1016/j.jneumeth.2018.03.010.

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

Peeters, A., Segundo-Ortin, M., 2019. Misplacing memories? An enactive approach to the virtual memory palace. Conscious. Cogn. 76, 102834 https://doi.org/10.1016/j.concog.2019.102834.

Pflueger, M.O., Stieglitz, R.D., Lemoine, P., Leyhe, T., 2018. Ecologically relevant episodic memory assessment indicates an attenuated age-related memory loss - a virtual reality study. Neuropsychology 36, 680–689.

Plechatá, A., Nekovářová, T., Fajnerová, I., 2021. What is the future for immersive virtual reality in memory rehabilitation? A systematic review. NeuroRehabilitation 1–24. https://doi.org/10.3233/nre-201534. Preprint.

Reggente, N., Essoe, J.K.Y., Baek, H.Y., Rissman, J., 2020. The method of loci in virtual reality: explicit binding of objects to spatial contexts enhances subsequent memory recall. J. Cogn. Enhanc. 4, 12–30. https://doi.org/10.1007/s41465-019-00141-8.

Reggente, N., Essoe, J.K.Y., Baek, H.Y., Rissman, J., 2020. The method of loci in virtual reality: explicit binding of objects to spatial contexts enhances subsequent memory recall. J. Cogn. Enhanc. 4, 12–30. https://doi.org/10.1007/s41465-019-00141-8.

Robin, J., Buchsbaum, B.R., Moscovitch, M., 2018. The primacy of spatial context in the neural representation of events. J. Neurosci. 38, 2755–2765. https://doi.org/10.1523/JNEUROSCI.1638-17.2018.

Rouaud, O., Graule-Petot, A., Couvreur, G., Contegal, F., Osseby, G.V., Benatru, I., Giroud, M., Moreau, T., 2006. Apport de l'évaluation écologique des troubles exécutifs dans la sclérose en plaques. Rev. Neurol. (Paris) 162, 964–969. https://doi.org/10.1016/S0035-3787(06)75106-2.

Ruet, A., Brochet, B., 2020. Cognitive assessment in patients with multiple sclerosis: from neuropsychological batteries to ecological tools. Ann. Phys. Rehabil. Med. 63, 154–158. https://doi.org/10.1016/j.rehab.2018.01.006.

Thapa, N., Park, H.J., Yang, J.-G., Son, H., Jang, M., Lee, J., Kang, S.W., Park, K.W., Park, H., 2020. The effect of a virtual reality-based intervention program on cognition in older adults with mild cognitive impairment: a randomized control trial. J. Clin. Med. 9, 1283. https://doi.org/10.3390/jcm9051283.

Tomioka, K., Kurumatani, N., Hosoi, H., 2016. Association between social participation and instrumental activities of daily living among community-dwelling older adults. J. Epidemiol. 26, 553–561. https://doi.org/10.2188/jea.JE20150253.

Troyer, A.K., Leach, L., Strauss, E., 2006. Aging and response inhibition: normative data for the victoria stroop test. Aging Neuropsychol. Cogn. 13, 20–35. https://doi.org/10.1080/138255890968187.

Varela-Aldás, J., Buele, J., Lorente, P.R., García-Magariño, I., Palacios-Navarro, G., 2021. A virtual reality-based cognitive telerehabilitation system for use in the covid-19 pandemic. Sustainability 13, 1–24. https://doi.org/10.3390/su13042183.

Ventura, S., Brivio, E., Riva, G., Baños, R.M., 2019. Immersive versus non-immersive experience: exploring the feasibility of memory assessment through 360° technology. Front. Psychol. https://doi.org/10.3389/fpsyg.2019.02509.

von Stülpnagel, R., Steffens, M.C., 2012. Can active navigation be as good as driving? A comparison of spatial memory in drivers and backseat drivers. J. Exp. Psychol. 18, 162–177. https://doi.org/10.1037/a0027133.

Wallet, G., Sauzéon, H., Pala, P.A., Larrue, F., Zheng, X., N'Kaoua, B., 2011. Virtual/real transfer of spatial knowledge: benefit from visual fidelity provided in a virtual environment and impact of active navigation. Cyberpsychol. Behav. Soc. Netw. 14, 417–423. https://doi.org/10.1089/cyber.2009.0187.

Ward, G., Jagger, C., Harper, W., 1998. A review of instrumental ADL assessments for use with elderly people. Rev. Clin. Gerontol. 8, 65–71. https://doi.org/10.1017/S0959259898008089.

Wilson, P.N., Foreman, N., Gillett, R., Stanton, D., 1997. Active versus passive processing of spatial information in a computer-simulated environment. Ecol. Psychol. 9, 207–222. https://doi.org/10.1207/s15326969eco0903_3.