Virtual Reality (2024) 28:100
https://doi.org/10.1007/s10055-024-00999-w

ORIGINAL ARTICLE

Testing memory of a VR environment: comparison with the real environment and 2D pictures

Merylin Monaro[1] · Cristina Mazza[2] · Marco Colasanti[3] · Elena Colicino[4] · Francesca Bosco[5] · Eleonora Ricci[2] · Silvia Biondi[5] · Michela Rossi[5] · Paolo Roma[5]

Received: 26 April 2022 / Accepted: 9 April 2024
© The Author(s) 2024

Abstract
In recent years, there has been a growing trend in cognitive psychology research towards recreating experimental situations in virtual reality (VR). VR settings are thought to have higher ecological validity than laboratory settings using digital, two-dimensional (2D) pictures. Some studies have shown cognitive performance in VR settings to follow that of the real world. However, other studies obtained controversial results. The present study tested the memory performance of three groups of participants who were exposed to the same environment (a room) through different modalities: in real life, in VR, and through 2D pictures. The results highlighted that participants who were exposed to the target room in real life had an overall better memory performance, compared to participants who saw the room in VR or through 2D pictures. On the other hand, no differences in memory performance emerged between the VR and 2D picture groups, except for the non-suggestive verbal task. The results suggest that future research should be careful in assuming that performance in VR settings is comparable to real life and that VR is more ecological than traditional 2D media.

Keywords Visual memory · Suggestibility · Virtual reality · Virtual environments

1 Introduction

Over the past decade, the increasing availability and decreasing cost of virtual reality (VR) technology has led to a considerable rise in research exploring VR applications. VR has become an especially popular research modality in experimental psychology, due to the almost limitless possibilities for creating complex and realistic scenarios with a high degree of experimental control while, supposedly, providing higher ecological validity than laboratory settings. Indeed, research has assumed that VR and reality are somewhat comparable, as long as certain conditions are met. In the early implementation of VR research designs, it was hypothesized that the similarity of users' responses between real-world and VR environments would be proportional to the degree to which the VR setting simulated "naturalistic" experiences (Bell et al. 2001)—in other words, its degree of realism (i.e., how faithfully it represented real-world input on all sensory channels and the fidelity of its environmental responses; Freeman et al. 2000). To this end, research has highlighted the role of a series of constructs in determining users' perception of VR environments. Among these, two closely related concepts—presence and immersion—seem significant. Presence indicates the subjective "sense of being there" (Freeman et al. 1999), while immersion is related to the objective properties of a system (Slater 2009), in terms of its replication of vivid, multisensory perceptions. On this point, it should be noted that the umbrella term "VR"

Merylin Monaro and Cristina Mazza contributed equally to this work.

✉ Merylin Monaro
merylin.monaro@unipd.it

✉ Cristina Mazza
cristina.mazza@uniroma1.it

[1] Department of General Psychology, University of Padova, Padova, Italy

[2] Department of Neuroscience, Imaging and Clinical Sciences, University "G.d'Annunzio," Chieti-Pescara, Chieti, Italy

[3] Department of Psychological, Health and Territorial Sciences, G. d'Annunzio University of Chieti-Pescara, Chieti, Italy

[4] Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, USA

[5] Department of Human Neuroscience, Sapienza University of Rome, Rome, Italy

Published online: 23 April 2024

may apply to a variety of devices (e.g., flat screen three-dimensional [3D] environments, head-mounted displays [HDM], room-based systems such as the Cave Automatic Virtual Environment [CAVE]) offering different levels of immersion. It has been suggested that VR can elicit a sense of presence that is akin to being present in real life, and thus greater than the sense of presence that can be achieved through interaction with traditional 2D media (e.g., Wagler and Hanus 2018).

While the literature seems to agree on this high sense of subjective presence in VR settings, the implications of this are debated. For instance, some researchers (e.g., Hodges et al. 1994; North et al. 1998) have linked the sense of presence in VR settings with increased emotional arousal, but not to improved task performance. Mania and Chalmers (2001) compared recollection performance of a 15-minute seminar delivered across four different conditions (i.e., in person, on a 3D desktop, through a 3D HMD, via audio), finding that level of presence was not associated with accurate memory recall, and that recall was significantly higher in the real-life condition, compared to the VR condition. Similarly, Slater et al. (1996) found that immersion—but not presence—increased task performance, which involved comprehension and memory of a complex 3D object. However, other studies investigating performance differences between VR and real-life environments have found contradictory results. For instance, Hu-Au and Okita (2021), in a study assessing environmentally-related learning differences, found comparable learning of general content knowledge in VR and real-life conditions. Conversely, Taylor and Dando (2018) compared episodic retrieval performance during interviews in a virtual avatar-to-avatar environment (i.e., with both interviewer and interviewee represented by avatars) and a traditional face-to-face environment, finding that participants in the avatar-to-avatar interview had significantly better recall.

Research has also investigated user experience and performance differences between two-dimensional (2D) display environments and 3D virtual environments, finding that: at the lowest level, the main difference between these environments is that VR provides users with a sense of depth and proportion that is lacking in traditional 2D media; and at higher levels, VR generally induces a stronger sense of presence and engagement (e.g., Radianti et al. 2020). Indeed, the international literature highlights that VR movies generate different EEGs and greater emotional arousal in viewers compared to 2D movies (Tian & Whang, 2021). Similarly, men (but not women) find VR pornography more sexually arousing than its 2D counterpart (Elsey et al. 2019). Some authors (see, e.g., Elmquaddem 2019) advocate for the use of VR as a learning tool, positing that VR "can improve and facilitate learning, increase memory capacity and make better decisions while working in entertaining and stimulating conditions" (p. 237). There is some corroboration for this claim. For instance, Schöne et al. (2017) reported that participants who watched a motorcycle ride via VR not only rated their experience as more realistic but also performed twice as well in a memory task than participants who experienced the same motorcycle ride via 2D video. Likewise, Krokos et al. (2019) found superior memory recall with an HMD compared to a traditional 2D desktop computer. Similarly, Norman et al. (2020) found a greater skin conductance response (taken to indicate recognition) to a mock crime scene presented in VR compared to 2D. Surprisingly, most studies investigating recall following exposure in VR in comparison to other modalities have not examined suggestibility. From a forensic perspective, interrogative suggestibility (i.e., the extent to which, within a closed social interaction, messages communicated during formal questioning are accepted, with a subsequent effect on behavioural responses; Gudjonsson and Clark 1986) could be an interesting variable to consider, as insight into this factor might contribute to the development of ecological mock crime scenarios.

However, two main problems arise: on the one hand, VR results are not consistent, as there are reports of both similar and worse memory performance based on interaction with VR, compared to 2D media (e.g., Ernstsen et al. 2019; Makransky et al. 2019; Kisker et al. 2021). Furthermore, an important caveat of research applying VR to learning is that, for VR to be effective, it must leverage its unique advantages, which include both presence and immersivity and embodiment and agency (e.g., Johnson-Glenberg 2019; Johnson-Glenberg et al. 2021). Indeed, research has confirmed that the utility of VR is dependent on a high degree of presence, immersion, and interactivity (e.g., Sutcliffe et al. 2005).

In light of this observation and the inconsistent findings on the differences between VR versus other media, the present study aimed at gaining an understanding of how the modality through which stimuli are presented (i.e., 2D vs. VR vs. real life) impacts memory recollection and suggestibility. For this purpose, three groups of participants were exposed to, respectively, a room in real-life, the same room in VR, and 2D pictures of the same room captured from different angles. The following hypotheses were formulated: (a) participants in the VR condition would perform similarly to participants in real life condition on free recall, visual recognition, non-suggestive visual and verbal questions, and resistance to suggestibility (verbal and visual questions) tasks; and (b) participants in the VR condition would perform significantly better than participants in the 2D picture condition on free recall, visual recognition, non-suggestive

visual and verbal questions, and resistance to suggestibility (verbal and visual questions) tasks.

2 Materials and methods

2.1 Participants

A total of 123 participants were volunteers who responded to a social media advertisement or were located near Sapienza. The inclusion criteria were: (a) aged at least 18 years; and (b) excellent comprehension of the Italian language. Four participants (3.25%) were excluded due to set-up issues related to the Meta Quest 2 device used in the study, which invalidated the procedure. The final sample comprised 119 participants, of whom 62 were male (52.1%) and 57 were female (47.9%), aged 18–35 years ($M = 24.20$, $SD = 4.130$). The majority of the sample were students ($N = 77$, 64.7%), educated to a high school level ($N = 67$, 56.3%), Italian citizens ($N = 118$, 99.2%), living in central Italy ($N = 107$, 89.9%), and experiencing no visual impairments ($N = 61$, 51.3%). A post hoc power analysis was computed using G*Power (Faul et al. 2007): a sample size of 119 resulted to be sufficiently large to achieve a statistical power (1-β) of at least 0.90 in a testing involving three groups, given a significance level of 0.05 and a large effect size (0.40).

Participants were randomly assigned to three groups using the Excel RAND function, according to a manipulated variable (i.e., the modality in which they visited the target room; see the "Measures" section for detailed information):

- - *Group 1 (G1)* ($M_{age} = 26.53$, $SD = 3.602$) was composed of 40 participants who visited the target room in real life.
- - *Group 2 (G2)* ($M_{age} = 23.8$, $SD = 3.556$) was composed of 40 participants who visited the target room in VR, using a Meta Quest 2 device.
- - *Group 3 (G3)* ($M_{age} = 22.79$, $SD = 3.600$) was composed of 39 participants who observed 2D pictures of the target room (captured from different angles) on a computer.

The mean age was statistically different between the three groups ($F_{2,116} = 11.344$; $p < 0.001$).

Table 1 reports the descriptive statistics for all of the characteristics considered, for each group and for the entire sample.

2.2 Measures

The following measures and instruments were used:

2.2.1 Measures used in phase 1 (see "Experimental Procedure" section)

Sociodemographic Questionnaire. Participants were administered a questionnaire to collect personal sociodemographic information on biological sex, age, education, occupational status, region of residence, citizenship, medical diagnoses, visual impairments, and prior experience with VR.

Rivermead Behavioural Memory Test (RBMT-III), "Figure Recognition" Subtest. The Rivermead Behavioural Memory Test (Wilson et al. 1985; Italian validation: Beschin et al. 2013) is an ecological assessment instrument that evaluates respondents' ability to use memory in everyday situations. Showing good ecological validity, it has great value in predicting real-life behavior and deficits outside the evaluation situation. The measure is composed of 14 subtests, aimed at evaluating visual memory, verbal memory, and recall memory aspects, both immediate and delayed. The present study administered the "Figure Recognition" subtest, which aims at testing respondents' ability to recall previously displayed images from a larger set.

Corsi Block-Tapping Test. The Corsi block-tapping test (De Renzi and Nichelli 1975; Spinnler and Tognoni 1987) is one of the most popular and widely used tests for measuring the quantity of information that can be held in short-term memory, otherwise known as visuospatial memory span. The stimulus is a board (32 × 25 cm) on which nine black cubes (4.5 × 4.5 × 4.5 cm) are attached asymmetrically. The cubes are progressively numbered on the face displayed to the examiner, who sits opposite the participant. The examiner taps the cubes in a prearranged sequence of increasing length (tapping a cube every 2 s). Immediately after the examiner finishes the sequence, the participant is asked to reproduce it, touching the cubes in the same order. The length of the sequence varies from 3 (the shortest) to 10 (the longest), and for each length, there are two prearranged sequences. If the participant correctly reproduces one of the sequences shown, the examiner progresses to a longer sequence. The participant's visuospatial memory span is reflected by the number of cubes related to the longest series correctly reproduced. The average visuo-spatial memory span is five (Spinnler and Tognoni 1987).

2.2.2 Measures used in phase 3 (see "Experimental Procedure" section)

Free Recall Task. Participants were asked to write down all of the objects they remembered seeing in the target room. Specifically, the instructions were: *"You have just seen a room. Please list in writing (without describing) all of the objects in the room. Let the examiner know when you have*

Table 1 Descriptive statistics of the sample and each group

	G1	G2	G3	TOTAL	χ^2	*p*
	***n*(%)**	***n*(%)**	*n* (%)	*N* (%)		
Biological sex					1.233	.540
Female	22 (55.0)	18 (45.0)	17 (43.6)	57 (47.9)		
Male	18 (45.0)	22 (55.0)	22 (56.4)	62 (52.1)		
Citizenship					1.992	.369
Italian	39 (97.5)	40 (100)	39 (100)	118 (99.2)		
Other	1 (2.5)	0 (0)	0 (0)	1 (0.8)		
Region					2.646	.619
North	0 (0)	2 (5)	1 (2.6)	3 (2.5)		
Central	36 (90)	36 (90)	35 (89.7)	107 (89.9)		
South	4 (10)	2 (5)	3 (7.7)	9 (7.6)		
Educational level					26.169	**< .001**
Middle school diploma	1 (2.5)	0 (0)	0 (0)	1 (0.8)		
High school diploma	11 (27.5)	27 (67.5)	29 (74.4)	67 (56.3)		
Degree	24 (60)	13 (32.5)	6 (15.4)	43 (36.1)		
Post-graduate	4 (10)	0 (0)	4 (10.3)	8 (6.7)		
Occupation					15.31	**.004**
Unemployed	1 (2.5)	1 (2.5)	2 (5.1)	4 (3.4)		
Student	17 (42.5)	30 (75)	30(76.9)	77 (64.7)		
Employed	22 (55)	9 (22.5)	7 (17.9)	38 (31.9)		
Vision impairment					.666	.717
No	19 (47.5)	20 (50)	22 (56.4)	61 (51.3)		
Yes	21 (52.5)	20 (50)	17 (43.6)	58 (48.7)		
Use of glasses / contact lenses					.224	.894
No	23 (57.5)	21 (52.5)	22 (56.4)	66 (55.5)		
Yes	17 (42.5)	19 (47.5)	17 (43.6)	53 (44.5)		
Medical/psychological/psychiatric diagnosis					2.097	.351
No	38 (95)	40 (100)	37 (94.9)	115 (96.6)		
Yes	2 (5)	0 (0)	2 (5.1)	4 (3.4)		
Prior experience of VR					1.069	.586
No	30 (75)	26 (65)	26 (66.7)	82 (68.9)		
Yes	10 (25)	14 (35)	13 (33.3)	37 (31.1)		

finished". Subsequently, participants were given 1 point for each object (out of 50) they were able to recall. The free recall task total score was based on how many objects of the target room participants were able to recall. The 50 items are listed in the Supplementary Materials.

Ad-hoc Questionnaire. An ad-hoc questionnaire was created and administered through the online software Qualtrics. The questionnaire included 60 questions related to the target room, meant to detect visual recognition, non-suggestive verbal and visual questions, and suggestibility (through verbal and visual suggestive questions). The questionnaire comprised two sections:

Section 1, Visual Recognition Task. Participants were shown a picture of an object / furniture item / painting and asked if they previously saw it in the room. Ten pictures represented items that were actually in the target room, while an additional 10 pictures depicted items that were not in the target room. Specifically, the question was as follows: *"Was the object / furniture item / painting in the room?"* Participants were allowed to write their answer, and they did not necessarily have to answer "Yes" or "No." Fig. 1 displays two items included in the visual recognition task.

Section 2, Suggestibility Task. The suggestibility task of the present study mimicked the structure of the Gudjonsson Suggestibility Scale-2 (GSS-2; Gudjonsson 1997) – adapting it to the research purpose and including a visual task – and presented participants with suggestive and non-suggestive (verbal and visual) questions related to the target room. The GSS-2 is a tool designed to measure interrogative suggestibility, which represents a person's propensity to accept information communicated during formal questioning with a subsequent influence on their responses. Specifically, the suggestibility task employed in the present study comprised: (a) 10 non-suggestive verbal questions, (b) 5 non-suggestive visual questions; (c) 15 suggestive

Fig. 1 *Two Items from the Visual Recognition Task Note.* The image on the left was on the wall in the target room, while the image on the right was not present in the target room

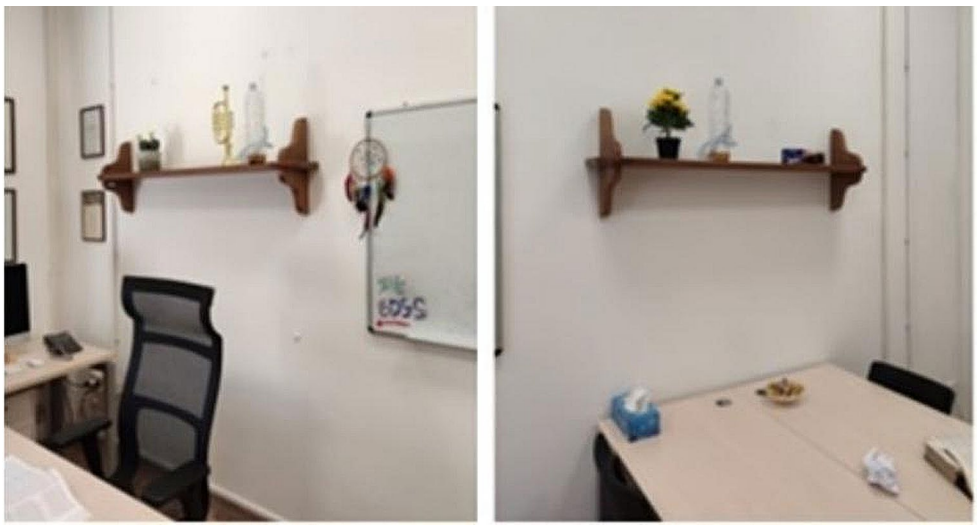

Fig. 2 One pair of alternatives presented in the non-suggestive visual section *Note*. the question related to this stimulus was: "was the bottle near the trumpet or near the flower pot?". note that the correct answer was "near the flower pot"

verbal questions; and (d) 10 suggestive visual questions. The non-suggestive verbal questions included five questions concerning true details of the target room (e.g., "*Was there a calendar in the closet?*") and five questions concerning false details (e.g., "*Was the mini fridge green?*"; note that there was a fridge in the room, but it was blue). The non-suggestive visual questions showed five pairs of photographs in which only one of each paired alternative represented reality (see Fig. 2 for an example question). The suggestive verbal questions asked about objects / furniture items / details that were not present in the target room (e.g., "*Was the backpack on the chair broken?*"; "*Was the carpet red or green?*"; note that there were no broken chairs or carpets in the room). Finally, the suggestive visual questions presented 10 pairs of photographs, both depicting in different locations an object that was not in the target room. Participants were, then, asked which of the two photographs represented the object's actual position, despite neither alternative was correct (see Fig. 3 for an example question). For this task and to allow participants to choose neither, the response mode was open (i.e., participants were not forced to answer "Yes" or "No" or "True" or "False", as in the GSS-2).

2.3 Experimental procedure

Data were collected in October 2021. The experimental procedure was conducted during working hours (9:00–17:00) to ensure adequate lighting conditions and took place in a neutral room and a target room of the Department of Human Neuroscience, "Sapienza" University of Rome. The experiment was designed in accordance with the Declaration of Helsinki and approved by the local ethics committee (Board of the Department of Human Neuroscience, Faculty of Medicine and Dentistry, Sapienza University of Rome). The experimental procedure lasted approximately 30 min and consisted of three phases: (1) assessment of participants' visual-spatial memory through neuropsychological tests, (2) exposure to the target room (real life vs. in VR vs. through 2D pictures), and (3) completion of the free recall, visual

Fig. 3 One pair of alternatives shown in the suggestive visual section *Note.* the question related to this stimulus was: "*was the backpack on the floor or on the chair?*" note that there was no backpack in the target room

Fig. 4 Picture of the target room

recognition, and suggestibility tasks in relation to the target room.

2.3.1 Phase 1

After providing written informed consent, participants completed the sociodemographic questionnaire and underwent a visual-spatial memory assessment through the Rivermead Behavioural Memory Test III "Figure Recognition" subtest and the Corsi block-tapping test (see the "Measures" section). These tests were useful to check the cognitive abilities of the participants, making sure that they had no visual-spatial memory deficits that could interfere with performance in the experimental task (phases 2 and 3).

2.3.2 Phase 2

Participants were randomly assigned to one of three experimental conditions: real life, VR, and 2D pictures.

Group 1 (G1): Real life Condition. Participants were taken into a target room of the Department of Human Neuroscience and positioned in the middle of the room. They were asked to memorize as many objects as possible over a period of 2 min, after which they were called by the experimenter. Participants had to stay in the middle of the room and could only rotate their body. Specifically, the instructions were as follows: *"We will now get you into a room. Your task is to observe the room carefully for 2 minutes, trying to memorize as many details as possible. We ask you to stand still at this point. You can turn your head in all directions and rotate around yourself."* Fig. 4 presents an image of the target room.

Group 2 (G2): VR Condition Using a Meta Quest 2. Participants were accompanied into a neutral room and asked to wear a Meta Quest 2 visor in order to visit the target room in VR. The Meta device has a 72 Hz LCD screen with a resolution of 1832×1920 pixels per eye. The visor is worn in front of the eyes and covers the entire field of vision. It also comprises two hand-held knobs that simulate hands. In the present study, only one knob was used, in order to permit participants to virtually access the room. Through the Meta Quest 2, participants were shown a panorama 360° picture of the target room, taken by a professional photographer with a Lapbano Pilot One EE. The 360° picture was taken from the same point where participants in Group 1 were standing.

The target room was the same as the room Group 1 explored in real life. After receiving guidance on the use of the Meta Quest 2 (e.g., that they should not move outside the planned area, and that they had to physically turn their head and body to see all parts of the room), participants were asked to memorize as many objects as possible over a period of 2 min, after which they removed the visor. Specifically, the instructions were as follows: *"Now we are going to give you a virtual reality experience. You will be in a room that you can visually explore in 360°. You will use this visor. Your task is to observe the room carefully for 2 minutes, trying to memorize as many details as possible. We ask you to stand still at this point. You can turn your head in all directions and rotate around."*

Following this step, participants were asked to answer two questions to assess their sense of presence inside the virtual environment. The first question (i.e., "*I felt completely immersed*") was adapted from Jennett et al.'s (2008) scale, as previously applied in other studies (e.g., Hudson et al. 2019); participants responded using a 7-point Likert scale ranging from 1 (*not at all*) to 7 (*very strongly*). The second question (i.e., "*I felt like I was inside the room*") was adapted from Wagler and Hanus's (2018) scale of spatial presence, and was measured on a 7-point Likert scale ranging from 1 (*not at all*) to 7 (*a lot*).

Fig. 5 Picture of the target room used in the 2D picture condition

Group 3 (G3): 2D picture condition. Participants were accompanied into a neutral room with a computer. They were asked to sit in front of the computer and look at some 2D pictures of the target room on the computer screen. Participants had 2 minutes to look at these pictures and memorize as many objects as possible. The eight pictures (2048 × 1537 pixels; see Supplementary Materials) showed different parts of the target room from the same point of view of participants in the real life and VR conditions. The pictures were sequentially shown on a 27" computer monitor, and participants could scroll across them as they wanted (see for example Fig. 5). Specifically, the instructions were as follows: *"Now you will be shown some pictures of a room. Your task is to observe the room carefully for 2 minutes, trying to memorize as many details as possible."*

2.3.3 Phase 3

After exposure to the target room, all participants were taken into a neutral room and administered the free recall task (see the "Free Recall Task" section).

Then, they completed an ad-hoc questionnaire about the room they observed on a 27" personal computer, with no time limit (see the "Ad-hoc Questionnaire" section). It was underlined that participants could answer according to their preference, and did not need to indicate "Yes" or "No."

3 Analysis and results

3.1 Data analysis

One-way independent ANOVA models were run to test performance differences between the three experimental groups (i.e., G1, G2, G3) in free recall, visual recognition, non-suggestive verbal questions, non suggestive visual questions, suggestive verbal questions, and suggestive visual questions. The effect sizes of the score differences between groups were reported; with respect to magnitude, $\eta^2 = 0.01$ was considered indicative of a small effect, $\eta^2 = 0.06$ a medium effect, and $\eta^2 = 0.14$ a large effect (Cohen 1988). To address the problem of multiple testing, Bonferroni correction was applied, dividing the p value by the number of tested variables ($n = 6$) and setting the significance level to 0.008 (Shaffer 1995). ANCOVA models were also run to test performance differences between the three experimental groups in free recall, visual recognition, verbal memory, visual memory, verbal suggestibility, and visual suggestibility; age, educational level and occupational status were entered as covariates, since these variables resulted statistically different between the three groups. Results are reported in Supplementary Materials.

As the three groups differed for age, educational level, and occupational status, to minimize any bias coming from the differences in covariates across groups, we employed two matching algorithms (i.e., the nearest neighbor matching and the coarsened exact matching) able to balance covariance discrepancies across groups through weights. We evaluated the performance of both algorithms and given the poor performance of nearest neighbor matching, we reported results using the coarsened exact matching (Iacus et al. 2012). The descriptive statistics and the density plots of the covariates before (pre-) and after (post-) matching procedure are reported in Supplementary Materials. We then used regressions of each outcome on the experimental condition and covariates including the matching weights to estimate the average effects of the experimental manipulation and tested the null hypothesis of no effect of the experimental manipulation. We included the covariates in the final regression as they can provide additional robustness to imbalances remaining after matching and can augment precision.

All analyses were performed using the SPSS v.28 software (IBM, 2021) and R (R Core Team 2021).

3.2 Results

3.2.1 Memory performance

Table 2 reports each group's average scores and standard deviations, and the ANOVA results. The ANOVAs generated significant results with respect to the free recall task, and the visual recognition task. Moreover, the ANOVAs indicated a significant effect of the experimental manipulation for the non-suggestive verbal questions, suggestive verbal questions, and suggestive visual questions. No significant results emerged from the ANOVA that explored differences between groups in relation to the non-suggestive visual questions. To address the problem of multiple testing, the Bonferroni correction was applied, dividing the p-value by the number of tested variables ($n = 6$) and setting the significance level to 0.008. Results from ANCOVAs are also reported in Supplementary Materials.

Results using the matching approach, which considers the differences in covariates across groups in age, educational level, and occupation, are mainly consistent with the ANOVAs results (without considering any covariate). Table 3 reports the difference in mean outcomes (average effect of the experimental condition) between the participants assigned to the different groups, after applying the coarsened exact matching. To address the problem of multiple testing, the Bonferroni correction was applied, dividing the p-value by the number of tested variables ($n = 6$) and setting the significance level to 0.008. The main results are the following.

Free Recall Task it emerged a statistically significant difference between G1 and G2 and between G1 and G3. In contrast, no differences emerged between G2 and G3. This indicates that participants who saw the room in real life had better recall of the room details compared to participants who explored the same room in VR or through 2D pictures.

Visual Recognition Task the analysis revealed a statistically significant difference between G1 and G2 and between G1 and G3. In other words, participants who were exposed to the room in real life performed better on the visual recognition task than participants who saw the same room in VR or through 2D pictures. No significant differences emerged between G2 and G3.

Suggestibility Task:

- ***Non-suggestive verbal questions***: there was a statistically significant difference between G1 and G3 and between G2 and G3. In contrast, no difference emerged between G1 and G2. These results indicate that participants who were exposed to the target room in real life had more accurate verbal recall than participants who were exposed to the same room through 2D pictures

Table 2 Average Scores (M) and Standard Deviations (SD) for Each Experimental Group (G1, G2, G3) on the Free Recall, Visual Recognition, Non-suggestive (Verbal And Visual) and Suggestive (Verbal And Visual) Tasks, and the Results of the One-Way Independent ANOVA Models (F-test, p-value, η^2)

	G1	G2	G3	F	*p*-value	η^2
	M (SD)	M (SD)	M (SD)	(df = 2,116)		
Free recall task	17.95 (3.75)	10.95 (3.04)	11.54 (3.48)	50.973	1.301e-16	0.468
Visual recognition task	17.18 (1.57)	14.18 (2.05)	13.80 (2.31)	34.239	2.06e-12	0.371
Non-suggestive questions						
Verbal	9.28 (0.82)	8.43 (1.06)	7.85 (1.51)	15.115	1.467e-6	0.207
Visual	4.70 (0.52)	4.48 (0.96)	4.13 (0.98)	4.587	0.012	0.073
Suggestive questions						
Verbal	1.75 (1.50)	3.05 (1.96)	4.08 (2.36)	13.916	3.828e-6	0.194
Visual	3.10 (2.59)	6.83 (2.85)	7.21 (2.52)	28.965	6.263e-11	0.333

Note G1: Real life Condition; G2: VR Condition; G3: 2D picture condition

Table 3 Difference in mean outcomes (average effect of the experimental condition) between the participants assigned to the different groups, after applying the coarsened exact matching

	Estimate	Pr(>|z|)	2.50%	97.50%
Free recall task				
G2 vs. G1	-5.9	< 0.001	-7.77	-4.04
G3 vs. G1	-5.71	< 0.001	-7.16	-4.26
G3 vs. G2	0.726	0.299	-0.643	2.09
Visual recognition task				
G2 vs. G1	-2.53	< 0.001	-3.29	-1.76
G3 vs. G1	-2.76	< 0.001	-3.19	-2.34
G3 vs. G2	0.0391	0.945	-1.08	1.16
Non-suggestive verbal questions				
G2 vs. G1	-0.661	0.0441	-1.31	-0.0176
G3 vs. G1	-1.32	< 0.001	-1.42	-1.21
G3 vs. G2	-0.491	0.00787	-0.853	-0.129
Non-suggestive visual questions				
G2 vs. G1	-0.409	0.189	-1.02	0.202
G3 vs. G1	-0.699	< 0.001	-0.86	-0.538
G3 vs. G2	-0.402	0.0509	-0.806	0.00166
Suggestive verbal questions				
G2 vs. G1	0.603	0.0568	-0.0175	1.22
G3 vs. G1	2.2	< 0.001	1.8	2.59
G3 vs. G2	0.834	0.0497	0.000931	1.67
Suggestive visual questions				
G2 vs. G1	3.09	< 0.001	1.7	4.48
G3 vs. G1	4.52	< 0.001	3.91	5.12
G3 vs. G2	-0.353	0.627	-1.78	1.07

but not than participants who were exposed to the same room in VR. Moreover, participants who were exposed to the target room in VR had a more accurate performance compared to those exposed to 2D pictures.

- ***Non-suggestive visual questions***: the analysis revealed a statistically significant difference between G1 and G3. This indicates that participants who were exposed to the target room in real life had more accurate visual recall than participants who were exposed to 2D pictures. No significant differences emerged between G1 and G2, and between G2 and G3.
- ***Suggestive verbal questions***: the analysis showed a statistically significant difference between G1 and G3 suggesting that participants in real life were significantly more resistant to verbal suggestions than those in the 2D condition. There was no significant difference between G2 and G1 and G2 and G3.
- ***Suggestive visual questions***: statistically significant differences emerged between G1 and G3, and between G1 and G2, whereas no difference emerged between G2 and G3, suggesting that participants in real life condition were significantly more resistant to visual suggestions than those in the other two experimental conditions.

3.2.2 Cognitive ability

To rule out the possibility that the three groups differed for cognitive ability, rather than the experimental condition, one-way independent ANOVAs were run to compare the performance of the three groups (i.e., G1, G2, G3) on the visual-spatial memory tests administered in Phase 1 of the experimental procedure (i.e., RMBT-III "Figure Recognition" subtest, Corsi block-tapping test). No significant results emerged for either the RMBT-III "Figure Recognition" subtest ($F_{(2,116)} = 0.948$, $p = 0.390$, $\eta^2 = 0.016$) or the Corsi block-tapping test ($F_{(2,116)} = 0.718$, $p = 0.490$, $\eta^2 = 0.012$), suggesting that there were no differences between groups in terms of basic memory skills (i.e., visual recognition, visual span).

3.2.3 Sense of presence

From the analysis of the questionnaire, it emerged that participants in the VR condition reported an appropriate sense of presence in response to both questions ("*I felt completely immersed*": $M = 5.48$; $SD = 1.43$; "*I felt like I was inside the room*": $M = 5.52$; $SD = 1.71$). A single sample *t*-test found that the means of both questions significantly differed from the central value of 4 (first question: $t_{30} = 5.759$, $p < 0.001$; second question: $t_{30} = 4.936$, $p < 0.001$).

4 Discussion

The present study aimed to examine how memory and suggestibility are affected by the media in which stimuli are presented. In more detail, three experimental groups were asked to memorize as many objects contained in a target room, shown respectively in real life, through a Meta Quest 2 HMD, and on a 2D desktop computer. Memory was assessed using free recall and visual recognition tasks, while suggestibility via verbal and visual tasks.

Compared to 2D, participants in the real life condition remembered significantly more details during free recall, made fewer errors in visual recognition and in both the non-suggestive verbal and visual tasks, and were more resistant to suggestive verbal and visual questions. These results highlight that viewing the stimuli in real life or in 2D might yield different performances in both memory and suggestibility tasks.

Similarly, compared to VR, participants in the real life condition remembered significantly more details during free recall and made fewer errors in visual recognition, hinting at the possibility that the memory performance in these two conditions is not comparable. Conversely, in relation to suggestibility, the performance between real life and VR did not significantly differed, except for the suggestive visual questions to which participants in real life were more resistant. While the impact of VR on suggestibility is still an underresearched topic, these results indicate that the ability to resist to suggestive questions might be somewhat similar in these conditions.

Additionally, VR participants obtained memory and suggestibility performances similar to those in the 2D condition, with the exception of making significantly fewer errors when answering non-suggestive verbal questions.

These results make a valuable contribution to the literature, emphasizing that users' performance in real life is not necessarily comparable to their performance in a VR setting, and that for a VR environment to elicit a lifelike response, it must do more than merely provoke a strong sense of presence (e.g., Mania and Chalmers 2001). The results are also partially aligned with previous studies finding no differences in performance between VR and 2D settings (e.g., Ernstsen et al. 2019; Makransky et al. 2019; Kisker et al. 2021).

At least two hypotheses could be formulated to explain why participants in the VR condition showed an overall worse performance on memory tasks than participants in the real life condition and similar performance to participants in the 2D condition. First, the VR stimuli used in the present study was a panorama 360° picture, rather than a computer-generated scenario; for this reason, although participants reported an appreciable sense of presence, other VR elements were missing, including interactivity and multisensoriality. Therefore, it may be the case that, for VR to be effective, it must leverage all of its unique advantages (i.e., embodiment and agency), as proposed by Johnson-Glenberg et al. (2021). However, the aim of the present study was to assess performance differences in memory tasks related to media presentation, net any other variable; thus, the stimuli presented in the three conditions were kept as similar as possible. Furthermore, the real life condition also lacked interactivity, as the participants were not allowed to freely explore the room.

A second hypothesis could explain the lack of performance differences on most tasks between the VR and 2D picture conditions: the literature indicates that memory performance improves when participants recall information in the same context in which the information was originally presented (e.g., Godden and Baddeley 1975). Therefore, considering that all participants completed the memory tasks on a computer, the feature similarity between the context in which participants in the 2D picture condition memorized the stimuli and carried out the memory tasks might have increased their performance to a level that was similar to that of participants in the VR condition. However, it should be also noted that participants in both the VR and real life conditions performed the memory tasks in a different environment than the one in which the information was learned.

There are three main limitations of the present study. First, considering that more than two-thirds of the sample had no prior experience with VR, more time could have been spent on the participants' training phase with the Meta Quest 2, in order to help participants become accustomed to the virtual environment. Second, as already mentioned, some advantages of VR (e.g., interactivity) were not leveraged, and this may have decreased the performance of participants in the VR condition. Finally, the third limit concerns the experimental stimulus, which consisted of a single item (i.e., the target room) shown in the three experimental conditions, thus the results should be interpreted with caution.

In conclusion, future research employing experimental paradigms in a VR environment should be careful in assuming that performance in a VR setting is comparable to performance in real life, and that VR environments are more ecological than traditional 2D media. Future research should also investigate the role of media characteristics on suggestibility. More research is needed to guide researchers in building VR environments with the aim of simulating real settings and measuring performance with high ecological validity. For example, future studies should investigate whether interactivity and multisensoriality are essential for VR environments to facilitate cognitive performance at real-life levels.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10055-024-00999-w.

Acknowledgements We thank Mr. Lucio Virzì, the photographer who took the 360° picture.

Author contributions CM, MM, and PR contributed to the study conception and design. MM and CM prepared the material. FB, MR, SB, ER, and MC collected the data. MM and EC performed the data analysis. All authors contributed to the data discussion and interpretation. CM, MM, MC, FB, and ER wrote the first draft of the manuscript. All authors commented on previous versions of the manuscript and read, revised, and approved the final version of the manuscript.

Funding Open access funding provided by Università degli Studi di Padova within the CRUI-CARE Agreement.

Data availability The datasets generated and/or analyzed for the present study are available from the corresponding author upon reasonable request.

Declarations

Competing interests The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Bell PA, Greene TC, Fisher JD, Baum A (2001) Environmental psychology, 5th edn. Harcourt College

Beschin N, Urbano T, Treccani B (2013) RBMT-3 Adattamento italiano. OS Organizzazioni Speciali Giunti Editore

Cohen J (1988) Statistical power analysis for the behavioral sciences. Routledge. https://doi.org/10.4324/9780203771587

R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

De Renzi E, Nichelli P (1975) Verbal and non-verbal short-term memory impairment following hemispheric damage. Cortex 11(4):341–354. https://doi.org/10.1016/s0010-9452(75)80026-8

Elmquaddem N (2019) Augmented reality and virtual reality in education. Myth or reality? Int J Emerg Technol Learn 14(3):234–242. https://doi.org/10.3991/ijet.v14i03.9289

Elsey JW, van Andel K, Kater RB, Reints IM, Spiering M (2019) The impact of virtual reality versus 2D pornography on sexual arousal and presence. Comput Hum Behav 97:35–43. https://doi.org/10.1016/j.chb.2019.02.031

Ernstsen J, Mallam SC, Nazir S (2019) Incidental memory recall in virtual reality: an empirical investigation. Proc Hum Factors Ergon Soc Annual Meeting 63(1):2277–2281. https://doi.org/10.1177/1071181319631411

Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39(2):175–191. https://doi.org/10.3758/bf03193146

Freeman J, Avons SE, Pearson DE, IJsselsteijn WA (1999) Effects of sensory information and prior experience on direct subjective ratings of presence. Presence: Teleoperators Virtual Environ 8(1):1–13. https://doi.org/10.1162/105474699566017

Freeman J, Avons SE, Meddis R, Pearson DE, IJsselsteijn W (2000) Using behavioral realism to estimate presence: a study of the utility of postural responses to motion stimuli. Presence: Teleoperators Virtual Environ 9(2):149–164. https://doi.org/10.1162/105474600566691

Godden DR, Baddeley AD (1975) Context-dependent memory in two natural environments: on land and underwater. Br J Psychol 66(3):325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Gudjonsson GH (1997) The Gudjonsson Suggestibility scales manual. Psychology

Gudjonsson GH, Clark NK (1986) Suggestibility in police interrogation: a social psychological model. Social Behav 1(2):83–104

Hodges L, Rothbaum BO, Kooper R, Opdyke D, Meyer T, de Graaf JJ, Williford JS (1994) *Presence as the defining factor in a VR application* (Technical Report No. GIT-GVU-94-5). Georgia Institute of Technology

Hu-Au E, Okita S (2021) Exploring differences in student learning and behavior between real-life and virtual reality chemistry laboratories. J Sci Edu Technol 30(6):862–876. https://doi.org/10.1007/s10956-021-09925-0

Hudson S, Matson-Barkat S, Pallamin N, Jegou G (2019) With or without you? Interaction and immersion in a virtual reality experience. J Bus Res 100:459–468. https://doi.org/10.1016/j.jbusres.2018.10.062

Iacus S, King G, Porro G (2012) Causal inference without balance checking: coarsened exact matching. Political Anal 20(1):1–24

IBM Corp (2021) IBM SPSS statistics for Windows, Version 28.0. IBM Corp, Armonk, NY

Jennett C, Cox AL, Cairns P, Dhoparee S, Epps A, Tijs T, Walton A (2008) Measuring and defining the experience of immersion in games. Int J Hum Comput Stud 66(9):641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004

Johnson-Glenberg MC (2019) The necessary nine: Design principles for embodied VR and active stem education. In P. Díaz, A. Ioannou, K. K. Bhagat, & J. M. Spector (Eds.), *Learning in a digital world: Perspective on interactive technologies for formal and informal education* (pp. 83–112). Springer Singapore. https://doi.org/10.1007/978-981-13-8265-9_5

Johnson-Glenberg MC, Bartolomea H, Kalina E (2021) Platform is not destiny: embodied learning effects comparing 2D desktop to 3D virtual reality STEM experiences. J Comput Assist Learn 37(5):1263–1284. https://doi.org/10.1111/jcal.12567

Kisker J, Gruber T, Schöne B (2021) Virtual reality experiences promote autobiographical retrieval mechanisms: electrophysiological correlates of laboratory and virtual experiences. Psychol Res 85(7):2485–2501. https://doi.org/10.1007/s00426-020-01417-x

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

Makransky G, Terkildsen TS, Mayer RE (2019) Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learn Instruction 60:225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Mania K, Chalmers A (2001) The effects of levels of immersion on memory and presence in virtual environments: a reality centered approach. Cyberpsychology Behav 4(2):247–264. https://doi.org/10.1089/109493101300117938

Norman DG, Wade KA, Williams MA, Watson DG (2020) Caught virtually lying: crime scenes in virtual reality help to expose suspects' concealed recognition. J Appl Res Memory Cognition 9(1):118–127. https://doi.org/10.1016/j.jarmac.2019.12.008

North MM, North SM, Coble JR (1998) Virtual reality therapy: an effective treatment for phobias. Stud Health Technol Inform 58:112–119

Radianti J, Majchrzak TA, Fromm J, Wohlgenannt I (2020) A systematic review of immersive virtual reality applications for higher education: design elements, lessons learned, and research agenda. Comput Educ 147:103778. https://doi.org/10.1016/j.compedu.2019.103778

Shaffer JP (1995) Multiple hypothesis testing. Ann Rev Psychol 46(1):561–584. https://www.annualreviews.org/doi/https://doi.org/10.1146/annurev.ps.46.020195.003021

Slater M (2009) Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos Trans R Soc Lond B Biol Sci 364(1535):3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater M, Linakis V, Usoh M, Kooper R (1996) Immersion, presence and performance in virtual environments. Proc ACM Symp Virtual Real Softw Technol – VRST '96. https://doi.org/10.1145/3304181.3304216

Spinnler H, Tognoni G (1987) Standardizzazione E taratura italiana di test neuropsicologici. Ital J Neurol Sci 8:1–113

Sutcliffe A, Gault B, Shin JE (2005) Presence, memory and interaction in virtual environments. Int J Hum Comput Stud 62(3):307–327. https://doi.org/10.1016/j.ijhcs.2004.11.010

Taylor DA, Dando CJ (2018) Eyewitness memory in face-to-face and immersive avatar-to-avatar contexts. Front Psychol 9:507. https://doi.org/10.3389/fpsyg.2018.00507

Wagler A, Hanus MD (2018) Comparing virtual reality tourism to real-life experience: effects of presence and engagement on attitude and enjoyment. Communication Res Rep 35(5):456–464. https://doi.org/10.1080/08824096.2018.1525350

Wilson BA, Cockburn J, Baddeley A (1985) Rivermead behavioural memory test. Thames Valley Test Company

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.