How individual factors intervene in the expression of cognitive abilities in a virtual immersive task: a validation study of the Virtual Reality Action Test

Tommaso Palombi  
“Sapienza” University of Rome

Federica Galli  
Foro Italico University of Rome

Francesco Giancamilli  
“Sapienza” University of Rome

Monica D’Amico  
“Sapienza” University of Rome

Fabio Alivernini  
“Sapienza” University of Rome

Luigi Gallo  
National Research Council

Pietro Neroni  
National Research Council

Marco Predazzi  
Fondazione Il Melo Onlus

Giuseppe De Pietro  
National Research Council

Fabio Lucidi  
“Sapienza” University of Rome

Antonio Giordano  
Temple University

Andrea Chirico (✉️ andrea.chirico@uniroma1.it)  
“Sapienza” University of Rome

Keywords: Virtual reality, performance-based test, sense of presence, immersion, cognitive assessment

Posted Date: March 8th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2598757/v1

License: ☑️ ☐️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on August 17th, 2023. See the published version at https://doi.org/10.1038/s41598-023-40510-0.
Abstract
There is a raised interest in literature to use Virtual Reality (VR) technology as an assessment tool for cognitive domains. One of the essential advantages of transforming tests in an immersive virtual environment is the possibility of automatically calculating the test’s score, a time-consuming process under natural conditions. Although the characteristics of VR can deliver different degrees of immersion in a virtual environment, the sense of presence could jeopardize the evolution of these practices. The sense of presence results from a complex interaction between human factors and the proprieties of VR devices. The present study has two aims: firstly, it contributes to the validation of a virtual version of the Naturalistic Action Test (i.e., Virtual Reality Action Test); second, it aims to evaluate the role of sense of presence as a critical booster of the expression of cognitive abilities during virtual reality tasks. The study relies on healthy adults tested in virtual and real conditions in a cross-over research design. The study’s results support the validity of the Virtual Reality Action Test. Furthermore, two structural equation models are tested to comprehend the role of sense of presence as a moderator in the relationship between cognitive abilities and virtual task performance.

1. Introduction
In recent years, the use of Virtual Reality (VR) technology and immersive interface have attracted the interest of several scholars. VR has been traditionally defined as “interactive, virtual image displays enhanced by special processing and by non-visual display modalities […] to convince users that they are immersed in a synthetic space.” The primary purpose of this technology is to provide an authentic and immersive experience, replacing real stimuli with high realism that can be customized and integrated into an ecological task. The simulated environment can be easily manipulated, facilitating experimental tasks that are difficult to implement in real-world settings. Other benefits of VR regard the possibility of participating in potentially dangerous tasks, such as moving in a complex environment or applying psychological treatment to problems arising from phobias in a controlled ecological setting. Although VR technology has an interdisciplinary nature, in the last few years, it might appear that most developments in VR studies have focused on clinical aspects. Nowadays, VR emerges as a promising valuable technology as an effective medium for administering different interventions in the healthcare contexts. Moreover, VR is considered a valuable tool that could improve and automatize the processes of administering and scoring traditional performance-based assessments without jeopardizing ecological validity. Different research highlighted that the score calculated from VR devices reduces the number of errors that could distort the results and reduces the time required for the scoring. The technological progress of VR has expanded the range of tools and types of research questions, adapting several standard performance-based tests to the virtual version. Different studies applied VR tools to assess executive functions and spatial abilities, showing promising results. However, several performance-based tests are traditionally delivered in real-life settings (e.g., Executive Functions Performance test - EFPT; the Multiple Errands Test - MET; the Radial Arm Maze task – RAM) this practice was often revealed time-consuming and not always feasible.

One of the most widely used performance-based tests is the Naturalistic Action Test (NAT), a valid and reliable test that measures the functional abilities of everyday activities in order to assess any inefficiencies (e.g., mis-reaching an object). The NAT is sensitive to mild functional changes associated with cognitive aging and includes tasks of increasing complexity, such as lunch preparation. The NAT scoring system evaluates both the steps performed for the task and the errors identified in the action. Recently, Giovannetti and colleagues developed a non-immersive digitalized version of NAT designed with a touchscreen interface in which the scoring was automatically calculated by the software without the need for human coding. In a pilot study, conducted by the authors, the results showed a high sensitivity of the test in terms of predicting significant group differences in mild functional difficulties. Non-immersive systems include the development of a 2D virtual environment projected on a desktop to reproduce images of the world. In particular, immersion involves stimulating senses, interactions, and reality’s similarity to the stimuli used in virtual environments. This feature can depend on the properties of the technological system used to isolate the user from reality. Although the characteristics of VR technology can deliver different degrees of immersion in a virtual environment, the sense of presence experienced by participants is the result of a complex interaction between human factors and the proprieties of the VR devices. The sense of presence plays an important role in the VR experience, and it has been noted that it is often complicated to find a direct influence on performance in VR tasks. The scientific literature makes no definitive claims about the relationship between presence and performance. Considering the above, evaluating the sense of presence in the analysis of performances in virtual performance-based tests predicted by cognitive tests becomes important. Whether the sense of presence reported null or weak
effects on scores of performance-based tasks, it stands to reason that it could act as a moderator of the cognitive test on the scoring of a specific task delivered in a virtual environment.

In line with the above, the present study aimed to test a virtual immersive version of the NAT (i.e., Virtual Reality Action Test, VRAT), following the same protocol implemented by Chirico and colleagues (2020). Using Head-Mounted Display (HMD) devices, we administered the VRAT to a sample of healthy adults. Participants have been asked to perform the same task (i.e., breakfast) in a cross-over trial in both conditions: virtual vs real. Moreover, we administered a cognitive battery test to assess the participants’ cognitive function. Based on previous research, we hypothesized that cognitive tests could predict performance in VRAT. Moreover, we expected that the VR experienced by participants (i.e., sense of presence) could moderate the relationship between cognitive tests and VRAT performance. Finally, in line with the literature, we expected significant correlations between performances in both virtual and real tasks and cognitive tests.

2. Materials And Method

2.1 Participants

The study sample was composed of 16 adults. Exclusion criteria for both groups were as follows: non-Italian speaker, current or past neurological disorder or major medical illness (e.g., dementia, traumatic brain injury, schizophrenia, etc.), current psychiatric disorder (e.g., major depression), severe sensory or motor deficit that would preclude interaction with devices. Inclusion/exclusion criteria were evaluated via a self-report questionnaire at the time of recruitment and a brief interview following informed consent obtained from all participants. The study was approved by the Ethical Committee (Department of Psychology of Developmental and Socialization Processes at “Sapienza”, University of Rome). All methods were carried out in accordance with relevant guidelines and regulations and in accordance with the Declaration of Helsinki.

2.2 Procedure

All procedures were identical for all participants in a cross-over design and were completed in a single session. The procedure included the following steps: 1) written informed consent obtained by the participant, 2) training on the VRAT with controllers, 3) testing on the VRAT with controllers and its real version (order counterbalanced), 4) presence questionnaire (only following the VRAT), 5) cognitive tests and other questionnaires.

2.3 Performance-Based Functional Tests

The breakfast task was administered in a highly immersive virtual environment and in a real-world environment, following the administration of NAT. The breakfast task requires participants to prepare a slice of toast with butter and jelly and a cup of coffee with milk and sugar while seated at a table containing a toaster, two knives, one spoon, butter in a butter dish, sugar in a bowl, a bottle of milk, mug filled with warm water, bread, instant coffee, jelly jar, and a napkin at the central workspace. The shape of the table and the spatial arrangement of objects followed the procedures reported in the NAT manual (NAT Manual; mrri.org). The breakfast task was administered in real and virtual environments. In both conditions, participants were instructed to complete the task, as quickly as possible, and without making errors. They were asked to make their movements as clear as possible and to tell the examiner when the task is finished. Performance in real conditions was video-recorded for scoring.

2.4 Virtual Reality Action Test (VRAT)

The VRAT is an immersive VR task that includes an everyday task (i.e., preparing breakfast) designed to maximize ecological validity by simulating real kitchen and household objects. The VRAT environment includes accurate 3D models, spatial audio, and automatic and real-time collection of movement data. The VR system included the HTC Vive head-mounted display that provided a fully immersive experience in a virtual environment and the controllers that provided tactile feedback through vibration to enable interaction with virtual objects in the VRAT. The VRAT system runs on an MSI Trident Gaming Desktop with 8GB RAM and a GTX 1060 graphic card. The HTC Vive head-mounted display provides users with a fully immersive virtual environment. The system provides visual content through two OLED displays for a total resolution of 2160 x 1200 pixels with a 110-degree FoV and a frequency of 90 Hz. The equipment and software specifications were the same used in a recent case study conducted by Chirico and colleagues.
During the VRAT, the participants were in a seated position in front of a virtual desk with virtual objects. The controllers were used to interact with the VR environment in which the user's hand motions are directly mapped to the virtual hand movements. As soon as the virtual hand reaches an object, the object is highlighted to indicate that it is selectable and interactable. Users are instructed to press a trigger button to interact with virtual objects in VRAT and release the trigger to end the interaction. The first phase of the VRAT regarded a VR training session in order to familiarize participants with the virtual environment. The training included four mini-tasks that comprised elements of the breakfast task: (1) toast a slice of bread; (2) spread the jelly on toast; (3) add instant coffee to a cup; (4) add milk to a cup. The examiner controlled the presentation of each mini-task from a monitoring position and could correct errors in object identification or the performance of task steps. The participants were encouraged to ask any questions. After the VR training session, the participants completed a single test trial of the VRAT in which they have been asked to perform the breakfast task. Participants were instructed to complete the test trial in silence, as quickly as possible, without making errors, making their movements as clear as possible. At the end of the task, they were told to stop and declare the test over.

Although the VRAT includes the error monitoring module, performance quality, and accuracy on the real condition and virtual conditions were evaluated by three trained coders, who independently viewed and coded the recordings of the participants’ performances to ensure inter-rater agreement. The performance scores were composed by:

- **Total errors**: incorrect actions (commission), the failure to complete a step (omission), and off-task actions (additions).
- **Micro-errors**: inefficient but not overtly incorrect actions; this category of errors refers to subtle inefficient behaviors in performing the task (e.g., reaching an object not needed for the accomplishment of the step).
- **Accomplishment score**: an accomplishment point was assigned for each task step of the breakfast task completed without error (range = 0–16).

### 2.7 Virtual Reality measures

After the VRAT conditions, questionnaires about the experience were administered to the participants.

**Presence Questionnaire**

The participants were asked to complete the Italian version of the Presence Questionnaire. The questionnaire, in its original form, comprises 24 items that explore different aspects of the VR experience, rated on a 7-point Likert scale. The factors are distributed as follows: 7 items on Realism; 4 items on the Possibility to act; 3 items on the Quality of the interface; 3 items on the Possibility to examine; 2 items on Self-evaluation of performance. In this study, we used the following factors: Realism, Possibility to act, and Possibility to examine. Reliability was higher than 0.88 for each factor.

**Cybersickness Symptoms**

The Italian version of the VR Cybersickness Symptoms was proposed to the participants to evaluate two types of side effects caused by exposure to VR: visual effects, such as tired eyes, aching eyes, eyestrain, blurred vision, and difficulties focusing, and physical effects, general discomfort, fatigue, boredom, drowsiness, headache, dizziness, concentration difficulties, and nausea. Participants answered reporting on a 6-point Likert scale the presence of symptoms, higher scores indicate more severe symptoms. The original version of the scale was validated by Ames reporting that the maximum irrelevant difference was set at 0.2.

### 2.8 General screening, cognitive tests

A trained psychologist administered to the participants the Italian versions of several questionnaires. Questions about general health (e.g., presence of psychiatric, neurological, or oncological conditions; motor or visual problems) were collected to screen the presence of clinical conditions that may interfere with the study. Specific cognitive abilities were assessed as summarized in Table 1 Italian version of questionnaires were administered.
Table 1
Cognitive tests used within the study

<table>
<thead>
<tr>
<th>Variable</th>
<th>Test</th>
<th>Original Scale citation</th>
<th>Italian Scale used for the study</th>
<th>Validity/Reliability of the instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual memory</td>
<td>Brief Visual Memory Test Revised (BVMT)</td>
<td>Benedict et al., 1996</td>
<td>Argento et al., 2016</td>
<td>Test-retest reliability r = from 0.60 for Trial 1 to 0.84 for Trial 3</td>
</tr>
<tr>
<td>Verbal fluency</td>
<td>Category Fluency</td>
<td>Benton et al., 1984</td>
<td>Zarino et al., 2014</td>
<td>Test-retest reliability r &gt; 0.75</td>
</tr>
<tr>
<td>Processing speed</td>
<td>Trail Making Test (TMT) - Part B</td>
<td>Armitage 1946</td>
<td>Gaudino et al., 1995</td>
<td>Not reported</td>
</tr>
<tr>
<td>Working memory</td>
<td>Digit Span backward</td>
<td>Weiss et al., 2010</td>
<td>Monaco et al., 2013</td>
<td>Test reliability = 0.89</td>
</tr>
<tr>
<td>Processing speed and Visual perception</td>
<td>Symbol Search (SS)</td>
<td>Weiss et al., 2010</td>
<td>Orsini et al., 2013</td>
<td>Test reliability = 0.88</td>
</tr>
</tbody>
</table>

2.9 Data Analysis

Descriptive analyses were performed on the collected data. Cognitive test scores were also evaluated by calculating the standardized z-score for the participants considering the normative data. Three trained observers evaluated the real version of VRAT (Real Action Test; RAT) performance (i.e., accomplishment and errors). Where there was a different evaluation between the observers, an agreement was found after discussion according to the NAT manual.

Pearson correlations were performed to compute the bivariate correlations between the key variables of the study, using jamovi software.

In order to test our hypothesis, the models were analyzed by employing Variance-Based Structural Equation Modeling (VB-SEM; known as Partial Least Squares analysis), which was performed with the WARP PLS v.8.0 statistical software. In VB-SEM, measurement error is explicitly modeled through the construction of latent factors, much like a covariance-based SEM analysis. VB-SEM, on the other hand, estimates models using ranked data, which is distribution-free, unlike covariance-based SEM. Model complexity, sample size, and deviations of the variable distributions from normality have less impact on model estimation. According to published criteria for VB-SEM models, VB-SEM analysis can evaluate the model at the measurement and structural levels. At the measurement level, VB-SEM establishes construct validity of the latent factors using the average variance extracted (AVE) and the composite reliability coefficients ($\rho$), which should exceed 0.50 and 0.70, respectively. AVEs for latent variables support discriminant validity if their square roots exceed the correlation coefficient with other latent variables. At the structural level, VB-SEM estimates the overall adequacy of the set of hypothesized relations among the model constructs using the goodness-of-fit (GoF) index given by the square root of the product of the AVE and average $R^2$ for the model with values of 0.10, 0.25, and 0.36 correspond to small, medium, and large effect sizes for model fit, respectively. The average path coefficient (APC) and average $R^2$ (ARS) coefficients provide additional information about the model's adequacy, both of which should be statistically significantly different from zero. Furthermore, the full collinearity variance inflation factor (AFVIF) is used to measure the level of multicollinearity, with values lower than 3.300 indicating that there are no issues with multicollinearity. Missing data were imputed using linear regression interpolation as recommended. According to our hypothesis, we tested two moderation models, evaluating the role of sense of presence as a moderator in the relationship between 1) “Memory Moderation Model”: cognitive tests related to the memory (i.e., BVMT, Category fluency, and Digit span) domain and VRAT performance, and 2) “Processing Speed Moderation Model”: cognitive tests related to processing speed (i.e., TMT, SS) and VRAT performance. We adopted an alpha level of 0.05, however, given the small sample size we also reported a marginal effect considered as < 0.10.

3. Results

3.1 Characteristics of the sample
On average, participants were 50.1 years old ($SD = 2.2$; range = 20–80). The sample comprised 54% of females. According to the Jack & Bondi guidelines$^{52}$, only two participants showed a low cognitive impairment (for a complete characterization of participants, see supplemental material Table S1 at the following link: osf.io/rcmyq).

### 3.2 Cognitive test

Raw cognitive test scores along with age and education-adjusted normative-based z-scores are reported in Table S1 (osf.io/rcmyq). Scores on most tests of specific abilities fell within the average range, including tests of visual memory, verbal fluency, working memory, processing speed, and visual perception.

### 3.3 Correlations

The correlation matrix (Table 2) showed a strong correlation between VRAT Scores with RAT scores. Specifically, the VR accomplishment shared 55.5% of the variance with Real scores ($r = 0.745, p < 0.001$), on the same page, VR error and Real error shared 78.5% of the variance ($r = 0.89, p < 0.001$). Alongside these results, the relation between cognitive tests and both versions of the NAT showed a coherent picture, in which high performances were strongly related to accomplishment in a positive direction and negatively with errors. Specifically, the accomplishment score in VR showed a pattern of strong correlation ($r > 0.40$) with all of the cognitive tests, while the errors score in VR showed a stronger correlation ($r > 0.60$) with Symbol Search and TMT, and a lower correlation with memory tests (i.e., Digit span, Category fluency, BVMT). Considering the scores in real, the correlation matrix showed a similar pattern with VR scores, while a lower intensity for the relationship between accomplishment and cognitive tests than VR scores. The micro-errors in RAT correlated positively with errors in both conditions (i.e., VR and Real) and negatively with accomplishments in both conditions. Moreover, the micro-errors in RAT correlated with micro-errors in VRAT showing a Pearson index of 0.41. Concerning micro-errors in VRAT, the correlation matrix showed no significant relationships with the other parameters.

<table>
<thead>
<tr>
<th></th>
<th>BVMT</th>
<th>Category Fluency</th>
<th>Digit Span</th>
<th>Symbol Search</th>
<th>TMT</th>
<th>Acc VR</th>
<th>Error VR</th>
<th>Acc Real</th>
<th>Error Real</th>
<th>Micro-Error Real</th>
<th>Micro-Error VR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BVMT</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category Fluency</td>
<td>0.392</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Span</td>
<td>0.258</td>
<td>0.151</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symbol Search</td>
<td>0.368</td>
<td>0.492</td>
<td>0.410</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMT</td>
<td>0.425</td>
<td>0.450</td>
<td>0.166</td>
<td>0.524</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acc VR</td>
<td>0.429</td>
<td>0.560*</td>
<td>0.518*</td>
<td>0.565*</td>
<td>0.466</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error VR</td>
<td>-0.380</td>
<td>-0.430</td>
<td>-0.189</td>
<td>-0.624**</td>
<td>-0.698**</td>
<td>-0.685**</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acc Real</td>
<td>0.305</td>
<td>0.450</td>
<td>0.297</td>
<td>0.563*</td>
<td>0.142</td>
<td>0.745***</td>
<td>-0.663**</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error Real</td>
<td>-0.548*</td>
<td>-0.425</td>
<td>-0.300</td>
<td>-0.713**</td>
<td>-0.619*</td>
<td>-0.581*</td>
<td>0.886***</td>
<td>-0.712**</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micro-Error Real</td>
<td>-0.109</td>
<td>-0.112</td>
<td>0.210</td>
<td>-0.333</td>
<td>0.140</td>
<td>-0.217</td>
<td>0.479</td>
<td>-0.446</td>
<td>0.384</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Micro-Error VR</td>
<td>-0.017</td>
<td>0.126</td>
<td>0.324</td>
<td>0.239</td>
<td>0.142</td>
<td>0.150</td>
<td>0.105</td>
<td>0.180</td>
<td>0.016</td>
<td>0.416</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note.* BVMT = Brief Visual Memory Test; TMT = Trial Making Test; Acc VR = Accomplishment in Virtual Reality; Error VR = Error in Virtual Reality.

Acc Real = Accomplishment in Real condition; Error Real = Error in Real condition.
3.4 Moderation models

According to our hypothesis, the moderation models exhibited acceptable fit: “Memory Moderation Model” (GoF = 0.868; APC = 0.277, \( p < .05 \); ARS = .783, \( p < .001 \); AVIF = 1.579) and “Processing Speed Moderation Model” (GoF = .786; APC = .422, \( p = .01 \); ARS = .649, \( p < .001 \); AVIF = 3.483).

The “Memory Moderation Model” showed that the effects of BVMT on the VRAT score were marginally moderated (\( p = .07 \)) by the sense of presence, positively with accomplishments and negatively with errors. All the paths between cognitive test and VRAT performance were significant or marginally significant, and coherent in the direction.

The “Processing Speed Moderation Model” showed that the statistically significant effect of the Symbol search test on VRAT errors was significantly and negatively moderated by the sense of presence. Concerning the direct effects of the cognitive variables on the VRAT performances, the Symbol Search was significantly and positively related with accomplishments, and negatively with errors, while TMT showed a similar pathway, significant only with the errors performance. The path coefficients and the \( p \) values of the models are reported in Figs. 1 and 2.

4. Discussion

The main purposes of the study were: first, to evaluate the validity of the Virtual immersive version of the NAT; second, to evaluate the effect of cognitive tests in predicting the VRAT performances, considering also the role of sense of presence as moderator.

In the scientific literature, other scholars digitalized the NAT using not immersive technology with good results. Following this mainstream, Chirico and colleagues (2020) developed a virtual immersive version of NAT (VRAT) implemented through Head-Mounted display devices, demonstrating the feasibility of the VRAT in a single case study. The results of the study supported the idea of implementing the VRAT even in people with minimal computer experience or with no prior VR exposure, without any significant risk. Furthermore, from a descriptive point of view, the results of both conditions virtual and real suggested similar outcomes. In line with these suggestions, the present study evaluated the concurrent validity of the VRAT and NAT in a cross-over design study, relying on a sample of 16 participants evaluated on memory and processing speed domains. Correlation analyses showed significant relations between VRAT scores (i.e., accomplishments and errors) and 1) RAT scores (i.e., accomplishments and errors), and 2) tests of cognitive abilities. These results suggested the potential validity of the immersive VRAT for function assessment, as the correlations of the scores (i.e., accomplishment and errors) between the different conditions (i.e., VR and real) showed a high Pearson index (> 0.70). This result is in line with previous evidence supporting the validity of the digitalized automated version of the NAT, where the correlation was 0.47. Regarding the micro-errors, the correlation matrix showed different patterns in the two conditions. Specifically, the micro-errors in RAT showed a coherent pattern with the other scores, although the sample size has influenced the \( p \)-values of the correlations. Indeed, the micro-errors correlated positively with errors and negatively with accomplishments in both real and virtual conditions. Moreover, the micro-errors in RAT correlated with micro-errors in VRAT. Concerning micro-errors in VRAT, the null relationships reported in the correlation matrix could be attributed to the difficulty of the software detecting micro-errors in a virtual environment. Indeed, the automated scoring was inaccurate in tracking subtle functional difficulties. Automatic scoring was shown to be overly sensitive to the hand movements of the participants, leading to inaccurate results. The progress of technology may improve the accuracy in detecting micro-errors, the authors of the present study are involved in a new protocol in order to fix this issue. Despite the difficulties mentioned above, the software was accurate in tracking errors and accomplishments, showing promising results.

Correlation analyses were performed between RAT scores and cognitive abilities, besides the relation between error and digit span, all variables resulted in statistically significant or marginally significant correlation. Although not all correlations were significant, given the Pearson indices, it is reasonable to think that the small sample size could have influenced these results. In any case, despite the small sample size, these results are in line with the literature, highlighting the validity of the RAT test.

As expected, also the correlation matrix between VRAT measures and tests of cognitive abilities followed the same pattern of the RAT’s correlation with cognitive tests, suggesting the accuracy of the automated score in detecting the accomplishments and errors during the task. These results are in line with a previous study related to the validation of a virtual non-immersive version of NAT.
Taken together, these results provided preliminary validity data for the VRAT and suggest the utility of the VRAT as an objective and efficient measure of functional difficulties. Virtual immersive technology allows to develop and implements immersive virtual scenarios in which, unlike non-immersive technology, the users could be able to experience the same feelings as the real scenarios. The role of VR experience is still debated in the scientific literature. Specifically, previous studies investigated the role of the sense of presence experienced by participants during the VR tasks, suggesting that it is often complicated to find a direct effect on scoring. For this reason, it becomes important to evaluate the sense of presence experienced by participants testing its role in moderating the effect of cognitive tests on VR tasks, when this can be applied.

To the best of our knowledge, this is the first study that evaluates the role of VR presence in moderating the relationship between cognitive abilities and the score of performance-based tests in virtual immersive environments. Our results showed that the sense of presence experienced by the participants during the task moderated the effects of BVMT on VRAT scores and the relationship between symbol search and errors. Although the relationships between sense of presence and performance were not found in previous studies or were found weak associations, our findings highlighted that the VR presence may have a moderation effect in these contexts. On the same page, the highest direct effects in the two models showed a pattern that emphasizes the role of visual components in our VR task. As matter of fact, our results highlight as both cognitive tests, related to the visuospatial domain (i.e., visuospatial memory; BVMT, and visuospatial attention; symbol search) had the highest effects on VRAT performances and were the only to being moderated.

Maneuerer and colleagues (2020) tested the effect of the sense of presence in predicting spatial cognition performance in an immersive virtual test (Rod-and-Frame Test; RFT). The authors found a positive effect of the sense of presence and other variables (e.g., cybersickness, game experience) on the participant's performance. Moreover, the scholars tested a mediation model without any significant results. However, they highlighted the role of human factors (e.g., sense of presence, cybersickness, game experience) in the spatial immersive virtual tasks. Following this claim, our study provides an empirical contribution to the role of sense of presence as a moderator, considering the validation of the VRAT test in performing daily living activities with no spatial performances involved. Our results showed that the sense of presence was not a predictor of VRAT performances. However, correlation analysis showed a null effect in the relation between sense of presence and cognitive domain and VRAT performances. To speculate, these results can be explained by the fact that given the different mechanisms involved in the NAT, where spatial performances are not involved to complete the task, the sense of presence becomes a crucial booster of expressing cognitive abilities (i.e., daily living activities) through VR devices. Moreover, participants reported no symptoms of cybersickness in VRAT suggesting the feasibility of the VR devices for all of them. The present study posed some important questions related to the complex interaction between individual factors and devices (e.g., sense of presence). Our results represented a first step in understanding the process behind the individual variations of sense of presence, without any manipulation of the immersive levels provided by the technology.

Future studies should take into account the interaction between human factors and VR systems, focusing on the relationships between several human factors (e.g., cybersickness, gender, VR experience) and different levels of immersive technology, maybe varying the immersive levels of the virtual environment. Although several studies mentioned above have evaluated the direct effect of human factors on performances in VR tasks, future researchers should consider these factors as a moderator component in the resolution of the task in a virtual environment (e.g., performance-based test in VR). Although the present study showed promising results, given the small sample of participants, broader studies are needed to confirm these findings.

**Declarations**

**Data availability statement**

Data are available at the following link: osf.io/rcmyq

**Author contributions**

All authors conceived and planned the experiments, contributed to the interpretation of the results, provided critical feedback, and helped to shape the research, analysis, and manuscript. AC, FeG, FrG, PN, and MP carried out the experiments. GDP, PN, and LG programmed the software. AC, TP, MDA, and FA managed and analyzed data. MDA and TP coded visual data. AC and TP took the lead in writing the manuscript. AC, TP, AG, and FL critically revised the manuscript and approved the final submitted version.
Acknowledgment

The research has been funded by the following organizations: Sapienza, University of Rome, Department of Psychology of Social and Developmental Processes; Sbarro Health Research Organization (www.shro.org), PA USA; PNRR, partnership 8, spoke 4 (Trajectories for active and healthy aging). We would like to thank Ms. Angela Romano for coordinating the data collection procedure and recruiting.

References

42. Armitage, S. G. An analysis of certain psychological tests used for the evaluation of brain injury. Psychol Monogr 60, i–48 (1946).


Figures
Figure 1. Estimates of the structural equation model involving memory domain. 
Note. BVMT = Brief Visual Memory Test; Accomplishment VR = Accomplishment in Virtual Reality; Error VR = Error in Virtual Reality. Dashed lines refer to nonsignificant path estimates. * p < .10, * * p < .05, ** p < .01, *** p < .001

See image above for figure legend.
Figure 2. Estimates of the structural equation model involving processing speed domain.
Note. TMT = Trial Making Test; Accomplishment VR = Accomplishment in Virtual Reality; Error VR = Error in Virtual Reality. Dashed lines refer to nonsignificant path estimates. * p < .05, ** p < .01, *** p < .001

Figure 2
See image above for figure legend.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- SupplementarymaterialsVRNAT2.xlsx