Towards a fusion of control and ecological validity in virtual reality based cognitive testing: pilot results from our QuestMarket game

Soma Zsebi[1,2], Cecilia Sik-Lanyi[3,4] & Renáta Cserjési[2]

[1]Doctoral School of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary | [2]Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary | [3]University of Pannonia, Egyetem u. 10, 8200 Veszprém, Hungary | [4]Hungarian Research Network, Piarista u. 4, H-1052, Budapest, Hungary

[1]zsebi.soma@ppk.elte.hu, [2]cserjesi.renata@ppk.elte.hu, [3]lanyi.cecilia@mik.uni-pannon.hu

DOI: 10.54337/aau.icdvrat26.36

Abstract

Virtual Reality (VR) devices have been extensively utilized in neuropsychological evaluation in recent years, however, their comparability to traditional neuropsychological tests is always compromised. One reason is researchers tend to utilize stimuli rich virtual environments which can alter performance. To overcome this problem, we created a VR based shopping game, QuestMarket, which employs a real-life mimicking setting but with very minimal environmental stimuli and compared its variables to six traditional neuropsychological tasks. Our study showed moderate comparability between certain cognitive tests and QuestMarket variables indicating that VR based cognitive tests may still measure different functions even in very plain environments.

1. Introduction

Virtual reality (VR) based cognitive testing represents a groundbreaking and promising approach to evaluating individuals' mental capabilities both in research and clinical settings (Pieri et al., 2023). The tool allows a more ecologically valid way of assessing people compared to classical paper-and-pencil based tasks whilst traditional tests tend to be better at maintaining control over the testing session (Kourtesis et al., 2021; Parsons, 2015). Despite its many potentials, though, the novelty of the device and methodology necessitates a deeper understanding of participants' user experience, enabling researchers to comprehensively grasp and customize VR software according to the needs of the participants. This evokes the question of how controlled testing scenarios and ecological validity can be brought closer together.

One key point in VR based cognitive testing is concrete and targeted measurements besides recording the performance indicators in the VR task. Researchers currently tend to employ questionnaires to evaluate participants' cybersickness symptoms, user experience and emotional states (Kourtesis et al., 2021; Kourtesis et al., 2023; Maeng et al., 2021) with a few of them applying biophysiological measurements (Lee et al., 2023; Xue et al., 2024; Zsebi et al., 2025) but only a handful of them making direct comparisons between classical and VR based neuropsychological performance. These approaches may serve as good additions to explaining contextual factors of cognitive performance in VR based settings, however, in order to have a more continuous nature

of scientific results and make studies more comparable to each other, traditional and modern methods must find a way to be comparable. However, considering the novel nature of VR based tests, more exploratory ways are needed to set the grounds for future studies as it is currently not fully clear which variables of complex and stimuli rich VR environments can be compared to (Bürkner et al., 2019). Unfortunately, traditional neuropsychological data also has its limitations with measurement methods being vulnerable to sample size, test environmental factors and individual physiological differences (Immanuel et al., 2023; Raghavendra, 2025).

In summary, VR based tests and traditional tests have a distinction between them with the former utilizing more ecological validity for less controlled processes and the former having more control but not real-life mimicking tests whatsoever. Moreover, these two distinct categories of tests hardly ever get compared and examined in the same continuum of professional perspective, practically creating two simultaneously existing but non-merging fields. In order to overcome this problem, we created a VR based cognitive test, named QuestMarket, in which participants must navigate around a shopping mall buying items from a shopping list. The shopping mall's environment is intentionally made very plain with as few environmental parts as possible to minimize any potential distractor effect they might cause. In this way, the software allowed us to directly examine and analyse the efficiency of VR based cognitive testing and compare it to six traditional neuropsychological tasks.

2. Methods

2.1. Participants

28 participants were enrolled in our study recruited from a Hungarian university's students between ages 18–35 (Mean: 22.66, SD: 3.035). Participants were free from all neurological and psychiatric disorders and had normal or corrected-to-normal vision.

2.2. Materials

2.2.1. Virtual Reality Device and Task

An Oculus Quest 2 headset was utilized for the task on which the software ran. The QuestMarket game was created in Unity and was started by the participants using the set's controller. The game consisted of a small shopping mall with a hallway in the middle of it (the player started here) and three shops opening from different parts of said hallway: a consumables store, a bathroom items store and a kitchen store. Players could open a shopping list using the controllers in which they could see the different items to be bought listed as well as their required quantity. Players were instructed to memorize the list and buy all items needed from memory, although they were also told they were allowed to open the list as many times as they required. Using either joystick's buttons, the player could get a red basket into their hands which they could use to put items in it. All shops consisted of plain blue shelves and brown wooden tables with no further decoration to them. The shops also all had a self-serving checkout machine players could use to scan and pay for their items. An item was only considered bought if the player successfully scanned it and pressed the button on the machine stating "Pay". Generally, all skins used on the environment and items were very plain and slightly cartoonish, with almost all items only consisting of a single colour (Figure 1 and 2) During playthrough, participants' gameplay was recorder using the Oculus device's built-in recorder and was later analysed by hand by the experimenters. Variables recorded during playthrough were total time required to finish the shopping spree, number of looks participants

took on the shopping list, number of items correctly bought from the list, number of items falsely bought from the list and number of items left out from the list.

Figure 1 and 2: Screenshots from the game. The left picture shows one of the shops with consumable items and the lower picture show the main hallway with the shopping list open.

2.2.2. Traditional Neuropsychological Tasks

The traditional neuropsychological testing consisted of six classical cognitive tests, namely the Corsi Block-Tapping Task, the Backward Corsi-Block Tapping Task, the Tower of Hanoi task, the Stroop task, the Iowa Gambling Task and the Rey Auditory Verbal Learning Task. All tasks except for the Rey task was administrated on PsyToolkit using a standard PC and a Chrome Web Browser. The Rey task was administrated by the experimenter using a printed version of the task. In both of the Corsi tasks, the length of the longest successfully recalled sequence was considered the final score; in the Tower of Hanoi task, the number of steps required to finish the task the score; in the Stroop task, participants faced 40 trials of different words and their final score was the number of correctly identified stimuli; in the Iowa Gambling Task, their final score was the final amount of money they finished the task with; and in the Rey task, their final score was the sum of all words recalled throughout all sessions.

2.3. Procedure

After filling out basic demographic data (sex, age, education level), participants either started with the neuropsychological testing first and then went on to finish the QuestMarket task or vice versa for which order they were randomly assigned to. Participants managed to do the QuestMarket task three times. First, they did it on easy mode for which they were told it was a practice run to get familiar and comfortable with the settings and the game. After that, they did the game once again on easy mode, this time, however, they were told it was an actual testing run. In reality, it was still a practice run and their performance was not recorded. We did this double practice run because in our pilot testing phases, we found that participants' performance shuffled due to the fact they thought it was a testing phase, likely due to the anxiety the testing caused. Thus, in order to get them used to being in an actual testing phase, we let them run the easy mode two times so that they could both get used to the controls first and then lower their test related anxiety level. Finally, after the two runs on easy mode, they went on to do the task on medium difficulty which was the actual, recorded run we later analysed. The game also consisted of a hard mode, however, in our pilot phases, we found that it was so hard, most performance scores appeared to be a floor effect for which reason we excluded it from our final procedure.

3. Results

All data were analysed using Jasp software 0.95.1. Considering the explorative nature of our study, we used Spearman's correlation for analysing potential connection between all variables. Ultimately, only two comparisons emerged as significant with a considerable effect size. The first one is between the Backwards Corsi Block-Tapping Task and the number of items correctly bought from the list positively correlated (rho=0.389, p=0.041) (Figure 3). Secondly, the Iowa Gambling Task negatively correlated with the number of shopping list looks (rho=-0.435, p=0.021) (Figure 3 and 4).

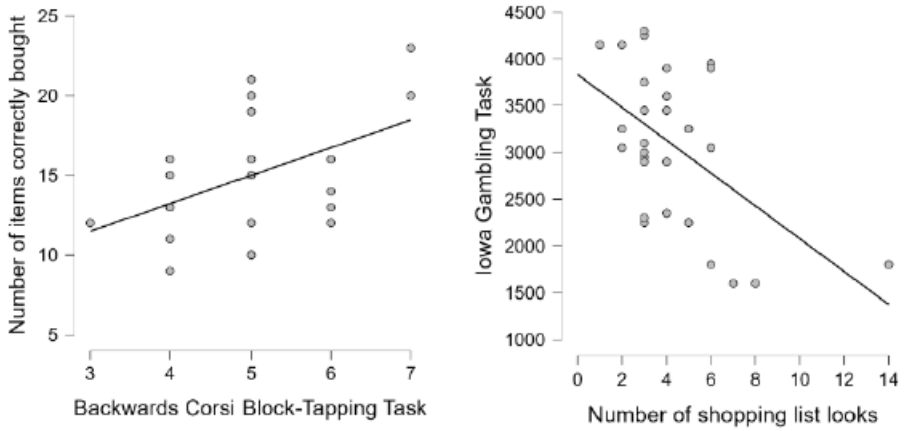

Figure 3 and 4: The upper figure shows Spearman's correlation between the score of the Backwards Corsi Block-Tapping Task and the number of items correctly bought during the QuestMarket task. The lower figure shows Spearman's correlation between the scores of the Iowa Gambling Task and number of shopping list looks during the QuestMarket task.

4. Discussion

In the present study, we investigated the comparability between VR based cognitive testing and traditional methods. In order to solely rely on the virtual task and minimize the effects virtual environments might evoke in participants, we created QuestMarket, a shopping game with a very plain and minimalistic virtual environment. Our results showed very little connection between QuestMarket variables and traditional cognitive test scores suggesting an altering effect for virtual environments even with very little environmental stimuli to influence participants. This gives rise to the potentiality that virtual environments fundamentally alter participants' performance on a variety of tasks. On the other hand, though, it can also mean that traditional neuropsychological tasks do not measure real-life comparable performance scores and their scores reflect very little of how a person would perform in a real life, cognitively demanding situation. Nevertheless, further studies are needed to confirm these initial assumptions, preferably with similar, plain virtual environments.

5. References

Bürkner, P. C., Schulte, N., & Holling, H. (2019). On the statistical and practical limitations of Thurstonian IRT models. *Educational and Psychological Measurement*, 79(5), 827–854.

Immanuel, S., Teferra, M. N., Baumert, M., & Bidargaddi, N. (2023). Heart rate variability for evaluating psychological stress changes in healthy adults: a scoping review. *Neuropsychobiology*, 82(4), 187–202.

Kourtesis, P., Collina, S., Doumas, L. A., & MacPherson, S. E. (2021). Validation of the Virtual Reality Everyday Assessment Lab (VR-EAL): An immersive virtual reality neuropsychological battery with enhanced ecological validity. *Journal of the International Neuropsychological Society*, 27(2), 181–196.

Kourtesis, P., Linnell, J., Amir, R., Argelaguet, F., & MacPherson, S. E. (2023, January). Cybersickness in virtual reality questionnaire (csq-vr): A validation and comparison against ssq and vrsq. In *Virtual Worlds* (Vol. 2, No. 1). MDPI.

Lee, J. Y., de Jong, N., Donkers, J., Jarodzka, H., & van Merriënboer, J. J. (2023). Measuring cognitive load in virtual reality training via pupillometry. *IEEE Transactions on Learning Technologies*, 17, 704–710.

Maeng, S., Hong, J. P., Kim, W. H., Kim, H., Cho, S. E., Kang, J. M., ... & Cho, S. J. (2021). Effects of virtual reality-based cognitive training in the elderly with and without mild cognitive impairment. *Psychiatry investigation*, 18(7), 619.

Parsons, T. D. (2015). Ecological validity in virtual reality-based neuropsychological assessment. In *Encyclopedia of Information Science and Technology, Third Edition* (pp. 1006–1015). IGI Global Scientific Publishing.

Pieri, L., Tosi, G., & Romano, D. (2023). Virtual reality technology in neuropsychological testing: A systematic review. *Journal of Neuropsychology*, 17(2), 382–399.

Raghavendra, S. (2025). Pupillometry as an Objective Measure of Auditory Perception and Listening Effort Across the Lifespan: A Review. *Journal of Otorhinolaryngology, Hearing and Balance Medicine*, 6(2), 24.

Xue, C., Li, A., Wu, R., Chai, J., Qiang, Y., Zhao, J., & Yang, Q. (2024). VRNPT: A neuropsychological test tool for diagnosing mild cognitive impairment using virtual reality and EEG signals. *International Journal of Human–Computer Interaction*, 40(20), 6268–6286.

Zsebi, S., Sik-Lanyi, C., Szabó, P., Filotás, P., Zsuffa, J. A., & Cserjési, R. (2025). A virtual reality take on the Corsi-Block Tapping Task: Exploring VR for spatial working memory assessment. *Acta Psychologica,* 261, 105918.