Reaction Time-Based Cognitive Assessments in Virtual Reality – A Feasibility Study with an Age Diverse Sample

Nils M. VAHLE[a,1], Sebastian UNGER[b], and Martin J. TOMASIK [a,c]

[a] *Developmental and Educational Psychology, Faculty of Health, Witten/Herdecke University, Witten, Germany*

[b] *Research Methods and Statistics, Faculty of Health, Witten/Herdecke University, Witten, Germany*

[c] *Institute for Educational Evaluation, University of Zurich, Zurich, Switzerland*

Abstract. Introduction: While virtual reality (VR) is an emerging paradigm in a variety of research contexts, VR-based embodiment effects on behavior and performance still lack in sufficient evidence regarding to bias in cognitive performance assessment. **Methods:** In this methodological observational study, we compare the VR measurement of cognitive performance with a conventional computer-based testing approach in real life (RL) in younger and older adults. The differences between VR and RL scenarios are investigated using the background of two theoretical models from cognitive psychology. Furthermore, data assessment reliability and validity are analyzed, concerning the feasibility of technological and ergonomic aspects. **Results:** A within-group comparison showed no change in information processing speed in either one of the two age groups, i.e., both groups perform equally well in RL and in a VR testing environment. **Conclusion:** The use of lifelike VR environments for cognitive performance tests seems not to lead to any performance changes compared to RL computer-based assessments, making VR suitable for similar applications. On technical concerns, we recommend the careful use of reaction time paradigms regarding to input hardware and stimuli presentation.

Keywords. Virtual reality, digital technology, feasibility studies, age groups, cognitive performance, psychology

1. Introduction

1.1. Scientific background

Experimental psychology has successfully utilized virtual reality (VR) environments to study avatar embodiment effects on affect, cognition, and behavior (e.g., concerning racial prejudice [1]). VR also offers a strongly immersive user experience, promising to maximize both internal and external validity. Medical applications show promising results with regard to the diagnostics and treatment of neurological disorders [2].

---

[1] Corresponding Author: Nils Vahle, Department of Psychology and Psychotherapy, University of Witten/Herdecke, Alfred-Herrhausen-Straße 50, 58448 Witten Germany; E-mail: nils.vahle@uni-wh.de.

However, when comparing VR to computer-based real life (RL) applications, the potential impact on humans' performance and affective state have not been studied in a systematic way. Furthermore, it remains unclear whether age-related performance declines known from conventional cognitive assessments [3] replicate to the same extent and with the same effect size under VR testing conditions.

The limited capacity model of motivated mediated message processing [4] implies that cognitive and perceptual processing speed as well as memory performance are directly connected to the complexity of stimuli. Information processing is seen as a combination of subprocesses that form a mental representation based on stimuli from several perceptual levels. Therefore, the combination of cognitive tasks with a novel and multi-sensory VR environment might pose an overload of this system and result in decreased performance.

Contrary to this line of reasoning, media research implies that stimuli from new, emotionally arousing multimedia content increase both attention and cognitive resources required for processing this content [5]. Thus, unknown or exciting content, such as seeing oneself as a digital avatar while using novel VR technology, might result in positive emotions such as excitement or joy and furthermore is expected to positively influence motivational activation and cognitive resource allocation.

1.2. Objectives of the study

The investigation focuses on the possible effects that a VR based assessment has on older and younger participants' cognitive performance in comparison to a RL assessment. Here, cognitive performance is tested by a reaction time-based assessment of perceptual speed, which declines in older age [6]. We aimed to carry reaction time tests to the extremes by including stimulus durations, that are nearly impossible to be properly processed by the human mind. Consequently, next to showing the limit of human perceptual speed, the feasibility threshold for VR-based cognitive performance tests is verified.

Due to the current literature, there is one first hypothesis: Cognitive performance declines in VR compared to RL, as stimuli from the VR environment strain cognitive capacity. The second hypothesis is that this effect is likely to be stronger in older participants, who seem more vulnerable to cognitive overload [7].

2. Methods

2.1. Study design

This 2×2 experimental study used a virtual reality intervention on two different age samples to observe performance within RL and VR test execution. Age group (older/younger) and test environment (RL/VR) served as independent variables.

The procedure started with (1) a demographic questionnaire. Then, (2) the information processing speed, which served as dependent variable, was tested in the RL environment, followed by mounting the VR headset. In a 90-second audio instruction, participants were asked to look carefully at their virtual surroundings and their virtual body. Afterwards the information processing speed test was repeated (3) in VR. The tests were carried out without any markable time in between in order to minimize possible bias, e.g., caused by mood on different days.

2.2. Setting

A VR programming studio created a laser-scanned digital copy of our laboratory room, including such details as furniture and lighting. Participants used an HTC Vive VR headset to enter the virtual lab room with an identical spatial orientation (Figure 1). A wall-mounted monitor presented the browser-based test and the corresponding stimuli within both environments, i.e., in RL over an Acer ET430K (43") and in VR over its digital copy. Participants were given the same handheld VR controller for all their tests to transmit click responses to the computer system. The study's headset is expected to minimize possible simulator sickness issues [8], especially in comparison to earlier hardware, due to fast and accurate movement tracking and 90 Hz display framerate.

Figure 1. Participant's view into the virtual laboratory room, showing the interactable monitor, the user's avatar and a fixed table in a mirror, and the user's floating left hand.

The study test was created via lab.js, a free, open-source experiment builder [9]. As its development environment has an easy-to-use visual interface to create a wide variety of studies, using four types of components, lab.js is more than suitable for this study. In addition, it can be used for in-laboratory and especially for online data collection [10].

2.3. Participants

In total, 58 participants (36 identified as female and 22 as male) were recruited via a digital bulletin board and a newspaper advertisement. Exclusion criteria were acute or chronical illnesses with regard to skeletal, neurological or mental health as well as pregnancy. Since the participants took part voluntarily and were not selected by anyone, there was no influence according to the age grouping process.

The participants formed two age-dependent groups, both consisting of females and males. There were 30 participants aged between 18 – 29 years ($M = 23.53$, $SD = 2.18$) in one group, while the other group comprised 28 persons aged between 50 – 78 years ($M = 62.29$, $SD = 5.69$).

2.4. Data measurement

The inspection time paradigm, which technical details were closely adapted to the original study [11], comprised the repeated on-screen presentation of basically four lab.js components. Their chronological sequence can be seen in Figure 2. Starting from left, a cross is used to focus the participants attention. It is followed by a simple shape with a

visual feature either on the left or on the right, that is selected randomly. Subjects were instructed to detect the side on which the feature appeared while all stimuli were presented with a randomized display duration (6 ms to 200 ms). Subsequently, it was briefly covered with a masking stimulus, which is also created with an image. The fourth component is a HTML page, containing two buttons. Participants responded via these buttons which side the feature was on.
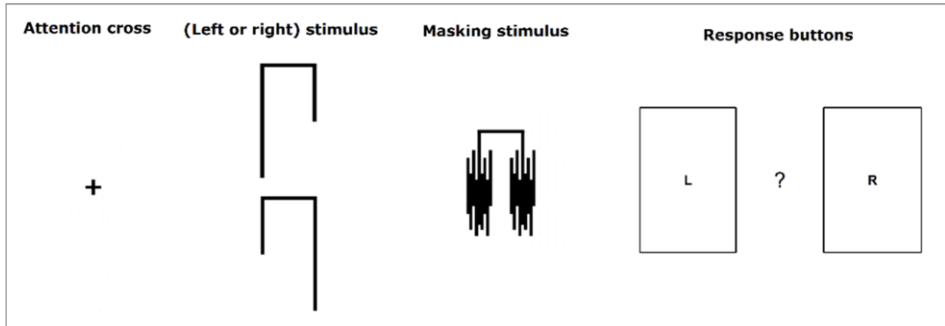

Figure 2. Visual components of the inspection time test. The chronological sequence starts on the left.

At the end of the test, there were a set of 60 responses for each participant with each of 15 display durations appearing 4 times. These responses were checked for correctness and then grouped by the 15 different display durations. Finally, accuracy scores were calculated, resulting in one mean score for each of the 15 display duration steps.

2.5. Data analysis

Data processing included the export of results at the end of each inspection time test and the calculation of their scores with the help of the environments Scilab and SPSS. A repeated measures ANOVA (analysis of variance [12]) was conducted as it allows the combination of a between-subject and a within-subject factor within the same analysis. The information processing speed data was compared between two test applications (RL vs. VR) and between two age groups (younger vs. older) using a single-factorial F-distributed function with $\alpha = 5\%$. The necessary assumptions were met with regard to dependence, scale level of variables, and absence of outliers. The assumptions of normal distribution and sphericity were not met, possibly due to the small sample size, but the Greenhouse-Geisser correction should keep the bias relatively low and the test conservative [13] to provide a robust result.

3. Results

As an attempt to exclude any learning effects, the accuracy scores were taken from the last 16 participants and equally divided into a first and second half for both environments. Weather within an environment ($\eta^2 = .02, p = .428$) nor after the switch ($\eta^2 < .001, p = .981$), such effect could not be found, suggesting a successful randomization of display durations.

For information processing speed, a significant effect of display duration was found in the whole sample ($\eta^2 = .611, p < .001$). With longer display durations, accuracy scores increased from the baseline guessing probability of 50 % to 100 %. Furthermore, there was a significant age effect ($\eta^2 = .193, p = .001$), i.e., younger participants performed better overall than older ones. Figure 3 displays the graphs of the accuracy scores separately for the age groups. The analysis also revealed an interaction effect when considering age group and display duration ($\eta^2 = .046, p = .007$), indicating the younger age group has a steeper slope in accuracy scores for increasing durations.
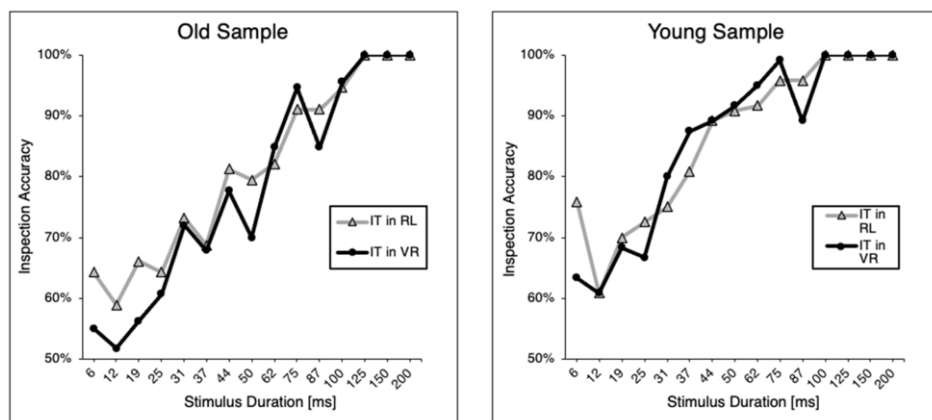

Figure 3. Inspection time (IT) accuracy scores of the old and young age group for both repetitions of the test (IT in RL and IT in VR) in accordance with the different stimulus durations between 6 and 200 ms.

Looking at the change in accuracy scores between the RL and VR of the total sample, there was no significance ($\eta^2 = .031, p = .186$). As a result, participants did not perform better or worse in one of the environments. In line with the study's hypothesis, data inspection at a descriptive level showed a moderate decrease in accuracy scores from RL to VR for the elderly only. Though, an interaction effect between age group and RL/VR assessment did not reach significance ($\eta^2 = .016, p = .347$), indicating a accuracy score stability for both age groups when switching from RL to the VR environment.

4. Discussion

4.1. Key results

This study investigated the effects of the human mind by first performing an assessment in RL and then in an identical VR environment. The effects of display duration are in line with similar test applications [14], indicating validity of our browser-based test application. The expected cognitive performance decline, resulting from an overload of information processing, was found neither in young nor in the elderly. However, this study demonstrates that the paradigm of the overall worse performance of older participants [14] can also be measured via VR without any loss of quality in the test. The same applies to the measurement of the inspection time paradigm [11] in general.

4.2. Limitations

Despite the careful implementation of the study, there are a few limitations and all conclusions have to be carefully reviewed against the backdrop of the experimental procedure. With the focus on feasibility, several hard and software components were implemented into the setup for the first time and especially the display duration of stimuli depends heavily on hardware and software characteristics. As an HTC Vive (frame rate of 90 Hz) refreshes its display after about 11.1 ms and an Acer ET430K (frame rate of 60 Hz) after only about 16.7 ms, the time durations of the original test could not be shown in appropriate length. Also, other time consuming computing processes cannot be excluded, so that participants could have recognized all stimuli clearly. An alternative could be a modified and validated inspection time test, e.g., with equally distanced stimulus durations (e.g., 25 ms, 50 ms, 75 ms, …). From a statistical perspective the low sample size might have led to delusion of the effects by error variance. Moreover, the two age groups were not equally in size. With the test setup starting with a RL assessment followed by VR, it is still possible that training effects systematically diluted the subsequent VR testing and resulted in performance increase. Thus, a truly valid evaluation of the VR effects would only be possible with an additional control group, in which the age groups complete the VR assessment prior to the RL assessment. Finally, this setup was mainly a visual task. Tasks with complex interactions of the surroundings should be investigated separately, as they require a more extrinsic cognitive load.

4.3. Conclusion

In summary, it has been shown that a standardized test for measuring human cognitive performance can be successfully implemented in a realistic VR setting. Based on the results, it can be assumed that VR is even equivalent to RL, making VR a promising tool for the diagnostics and treatment of various neurological or mental disorders. As it was a feasibility study, further investigations are recommended with regard to these results. An example for future applications is a follow up study, where an avatar age group specific performance difference on physical and cognitive performance domains was successfully induced [15]. Together with the present research, this is a crucial groundwork for applying the present and novel technique to the self-reflexive age related research, where participants embody a virtual avatar of a younger or older age.

By looking at the feasibility of data assessment in VR, this study discovered a technical limitation. An inspection of hardware characteristics and usability of peripherals showed that a setup with fewer interconnected devices for input and display might minimize aliasing effects and distortions of reaction time data. In addition, variations in auditory stimulus onset proved inaccuracy at the millisecond level. For future reaction time-based VR experiments, the use of tabletop response buttons is recommended. As long as certain technical limitations are considered, VR remains a promising tool for research, e.g., for the cognitive performance or the affective state.

Declarations

Ethical vote: Approved by the ethics committee at University Witten/Herdecke chaired by Prof. Dr. med. P. W. Haidzik, vote-no: *#216/2018*, December 5, 2018.

Study-Registry: not applicable

Conflict of Interest: The authors declare that there is no conflict of interest to disclose.

Author contributions: MT: conception of the work and study design; NV: data acquisition; NV, SU, MT: data analysis and interpretation; SU: Programming; NV, SU: writing the manuscript; NV, SU, MT: substantial revising of the manuscript. All authors approved the manuscript in the submitted version and take responsibility for the scientific integrity of the work.

References

[1]   Peña J, Hancock JT, Merola NA. The Priming Effects of Avatars in Virtual Settings. Communication Research 2009;36(6):838–56. doi: 10.1177/0093650209346802.

[2]   Garcia-Agundez A, Folkerts A-K, Konrad R, Caseman P, Göbel S, Kalbe E. PDDanceCity: An Exergame for Patients with Idiopathic Parkinson's Disease and Cognitive Impairment. In: Mensch und Computer 2017 - Tagungsband. Regensburg: Gesellschaft für Informatik e.V; 2017. p. 381–386.

[3]   Deary IJ, Corley J, Gow AJ, Harris SE, Houlihan LM, Marioni RE, et al. Age-associated cognitive decline. British Medical Bulletin 2009;92(1):135–152. https://doi.org/10.1093/bmb/ldp033.

[4]   Lang A. The Limited Capacity Model of Mediated Message Processing. Journal of Communication 2000;50(1):46–70. doi: 10.1111/j.1460-2466.2000.tb02833.x.

[5]   Lang A, Park B, Sanders-Jackson AN, Wilson BD, Wang Z. Cognition and Emotion in TV Message Processing: How Valence, Arousing Content, Structural Complexity, and Information Density Affect the Availability of Cognitive Resources. Media Psychology 2007;10(3):317–38.

[6]   Schaie KW. Perceptual speed in adulthood: Cross-sectional and longitudinal studies. Psychology and Aging 1989;4(4):443–453. https://doi.org/10.1037/0882-7974.4.4.443.

[7]   Malcolm BR, Foxe JJ, Butler JS, Sanctis P de. The aging brain shows less flexible reallocation of cognitive resources during dual-task walking: A mobile brain/body imaging (MoBI) study. Neuroimage 2015;117:230–42. doi: 10.1016/j.neuroimage.2015.05.028.

[8]   Singla A, Fremerey S, Robitza W, Raake A. Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays. In: 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX). Erfurt: IEEE; 2017. p. 1–6.

[9]   Henninger F, Shevchenko Y, Mertens UK, Kieslich PJ, Hilbig BE. lab.js: A free, open, online study builder; 2019.

[10] Bridges D, Pitiot A, MacAskill MR, Peirce JW. The timing mega-study: comparing a range of experiment generators, both lab-based and online. PeerJ 2020;8:e9414. doi: 10.7717/peerj.9414.

[11] Vickers D, Smith PL. The rationale for the inspection time index. Personality and Individual Differences 1986;7(5):609–23. doi: 10.1016/0191-8869(86)90030-9.

[12] Vincent WJ, Weir JP. Statistics in Kinesiology. 4th ed. Human Kinetics; 2012.

[13] Berkovits I, Hancock GR, Nevitt J. Bootstrap Resampling Approaches for Repeated Measure Designs: Relative Robustness to Sphericity and Normality Violations. Educational and Psychological Measurement. 2000;60(6):877-892. doi:10.1177/00131640021970961

[14] Ritchie SJ, Tucker-Drob EM, Deary IJ. A strong link between speed of visual discrimination and cognitive ageing. Curr Biol 2014;24(15):R681-3. doi: 10.1016/j.cub.2014.06.012.

[15] Vahle NM, Tomasik MJ. Declines in Memory and Physical Functioning when Young Adults Experience Being Old in Virtual Reality. Forthcoming 2021.