Intra-individual Stability and Assessment of the Affective State in a Virtual Laboratory Environment: A Feasibility Study

Nils M. Vahle[1], Sebastian Unger[2] and Martin J. Tomasik[3]
[1]*Developmental Psychology and Educational Psychology, Faculty of Health, Witten/Herdecke University, Witten, Germany*
[2]*Research Methods and Statistics, Faculty of Health, Witten/Herdecke University, Witten, Germany*
[3]*Institute of Education and Institute for Educational Evaluation, University of Zurich, Zurich, Switzerland*

Keywords: Virtual Reality, Human Computer Interaction, Age Paradigms, Affect, Experimental Psychology.

Abstract: While virtual reality (VR) emerges in a variety of research contexts, the effects on behavior and performance caused by VR-based embodiment still lack sufficient evidence of changes in affective state. With this feasibility study, we compared the affective states in both younger and older adults, measured after conventional computer-based tests in real life (RL) and after tests in VR. These assessment tests are spread over five time points, two in RL and three in VR, and the differences between the VR and the RL environment are investigated against the backdrop of two theoretical models of cognitive psychology. Results showed no change in affective state in either age group, switching from a RL to a VR environment. In addition, the elderly did not assess their affective state significantly different than that of the younger control group. In conclusion, lifelike VR environments for cognitive testing and other assessment or training purposes do not seem to lead to any systematic influence of affective state compared to RL computer-based assessments, making VR an alternative to conventional methods, for instance for cognitive treatments or preventions. Although the results can only be partially generalized due to a small sample size, they show technical stability and suitability for future use of similar applications.

1 INTRODUCTION

1.1 Background

In the past decade, attempts to improve human health with the help of virtual reality (VR) have intensified. For instance, VR-based physical exercises showed improved neuroplasticity compared to the traditional method, reducing the risk of mild cognitive impairment (Anderson-Hanley et al., 2012). Another study found that the topics and sessions of a diabetes education program in VR and real life (RL) were almost identical, but have their own advantages (Safaii et al., 2013). Even in the area of experimental psychology, VR successfully revealed effects of avatar embodiment on affect, cognition, and behavior, for instance with regard to racial prejudice (Peña et al., 2009). These examples show that VR can offer a promising immersive user experience with a wide range of uses. Nevertheless, the possible consequences and the impact on the affective state of the participants when comparing VR with RL were not further investigated.

Following the limited capacity model of motivated mediated message processing (Lang, 2000), we assumed that cognitive resources were limited and that the combination of cognitive tasks with a VR environment would lead to cognitive overload and thus to a stressful influence on the participant with effects on the affective state. Contrary findings from media research, however, would imply that stimuli from new, emotionally arousing multimedia content increase both the attention and the cognitive resources required and allocated for processing this content (Lang et al., 2007). Thus, the emotional reaction to novel or exciting content could be expected to influence motivational activation and cognitive resource allocation. Hence, whether using a VR testing environment will lead to the increase of negative or positive affect is yet unclear. Our research is meant to shed some light on this open question.

Vahle, N., Unger, S. and Tomasik, M.
Intra-individual Stability and Assessment of the Affective State in a Virtual Laboratory Environment: A Feasibility Study.
DOI: 10.5220/0010872200003123
In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 5: HEALTHINF*, pages 590-596
ISBN: 978-989-758-552-4; ISSN: 2184-4305

1.2 Objectives

This study investigated the effects of VR on affective state, using an age-diverse sample, i.e., younger and older participants. With the help of the following research question, VR and RL assessments were compared. We expected an increase in stressful arousal and a decrease in positive affect valence. The effects were expected to be stronger in older participants, who seem more vulnerable to cognitive overload (Malcolm et al., 2015). An investigation of a possible gender effect, which is reported by Goswami & Dutta (2015) as a significant variable in certain technology-related cases, is also of interest, but not as important as the age effect. Lastly, we aimed to explore the feasibility of cognitive performance testing in VR, focusing on hardware characteristics, usability of peripherals, data quality, and assessment reliability.

2 METHODS

2.1 Participants

Healthy adult participants were recruited over a digital bulletin board and a newspaper advertisement. Skeletal, neurological, or mental health illnesses as well as pregnancy led to exclusion from the study that was approved by the ethics committee at the Witten/Herdecke University in December 2018 (reference number 216/2018).

The study was part of a larger project on the virtual activation of age stereotypes, where a comparison of discrete age groups was planned. Recruiting younger participants (18 - 29) on campus as a young sample and older participants (50+) as relatives of students was an option chosen for pragmatic reasons after experiencing recruitment problems due to the COVID-19 pandemic. In total, there were 58 participants (36 female and 22 male), forming two age groups, regardless of gender. One group included 30 participants aged between 18 – 29 years (*M = 23.53*, *SD = 2.18*) and the other group comprised 28 persons aged 50 – 78 years (*M = 62.29*, *SD = 5.69*).

2.2 Setting

A special computer room was designed for the implementation of this study. Key objects included a table to position the participants, a mirror for self-observation, and a wall-mounted monitor to conduct various assessments. An exact copy of this room was created by a VR programming studio in which the spatial orientation remained the same. Using a laser scan, even small details, e.g., information signs, ventilation shafts, or light switches, could be transferred to the digital copy (Figure 1). The virtual laboratory room was “entered” via an HTC Vive VR headset and the interaction with the computer system took place via a handheld VR controller. To avoid any possible bias, the very same controller was used in RL and in VR.

Even though the main objective of this study was to test the feasibility of VR based assessments with regard to affective state differences compared to RL testing environments, participants were presented with a number of cognitive tasks. These tasks should firstly provide a certain quality of cognitive focus on details within the environment and secondly increase the time spent within the experimental VR setting.

Figure 1: Perspective to the operational monitor from RL (top) and VR (bottom).

All tests were browser-based and created via lab.js, a free, open-source experiment builder (Henninger et al., 2019). This builder has some advantages that meet the needs of this study. First, its development environment has an easy-to-use visual interface with a modular principle to create a wide variety of studies quickly. The components that the builder includes are HTML (Hypertext Markup Language) pages and images to design the frontend, loops to repeat an underlying component, and

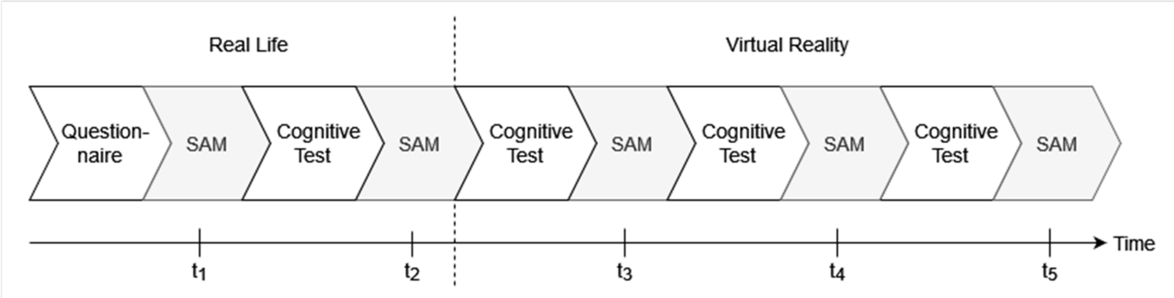

Figure 2: Procedure of stimuli and testing the affective states in chronological order.

sequences to create a chain of components. Second, tests created with lab.js can be customized with the languages HTML, CSS (Cascading Style Sheets), and JavaScript. Therefore, many custom modifications such as counting correct user inputs or including a beep were possible. Finally, lab.js has a suitable usage for in-laboratory and online data collection, while there is a low latency when measuring reaction times in the latter scenario (Bridges et al., 2020).

The study's procedure (see Figure 2) that was equal for all participants started with a demographic questionnaire. Then, the first affective state was assessed (t_1) followed by one of the three different cognitive performance tests (i.e., Inspection Time Test). After a second affective state assessment (t_2) in RL, the VR headset was mounted. Within this environment, all three cognitive tests (Inspection Time Test, Corsi Block Test, and Stop Signal Test) were presented in a randomized order. Each of the tests was followed by a further affective state assessment (t_3 / t_4 / t_5).

2.2.1 Inspection Time Test

As for all other tests, the lab.js builder replicated the Inspection Time Test. Here, a measurement was repeated overall 60 times to simulate the inspection time paradigm (Vickers & Smith, 1986). Starting with the actual chronological order of this test, there is a simple cross in the middle of the screen for 500 ms to get the participants' attention, which is followed by a simply shaped stimulus. This stimulus has a visual feature that randomly appeared either on the left or on the right. The participants were asked to determine the correct side of this feature within an also randomized display duration (6 ms to 200 ms), using one button as interactive input for each side. Before the stimulus completely disappeared, it was covered by a masking stimulus to disturb the short-term memory of the participants. In addition to the actual concentration test, there was an introduction and practice run, both in identical but simplified form.

2.2.2 Corsi Block Test

For testing the spatial working memory, the Corsi block-tapping test (Berch et al., 1998) was implemented. It was designed as near as possible after the original experimental setup. In this adaption, buttons spread across the screen, representing the nine rectangular shapes. These buttons were highlighted in black (750 ms) in a random order with a short pause in between (1 s). Then, participants were asked to replicate the order. An order increased in length (2 to 8) after each second trial when a participant replicated at least one of the trials correctly. Otherwise, the test ended in order not to discourage the participants.

2.2.3 Stop Signal Test

An adapted version of the stop signal test (Sahakian & Owen, 1992) presented randomly either a left- or a right-facing on-screen arrow next to two response buttons. Participants had to click the button that the arrow is pointing at as quickly as possible. After a training phase, requiring direction feedback only, the test was supported with an additional auditory stimulus. This loud beep featured a delay (250 ms to 2,000 ms) and was played while visualizing the arrow stimulus. Now, participants were asked to inhibit their feedback reaction if they heard the beep. This procedure was repeated 60 times, including 16 randomly distributed beep events.

2.3 Affective State

Our adaptation of the self-assessment manikin (SAM) was based on the version of Bradley and Lang (1994) and offered a visual on-screen visual scale with a semantic differential as it is reliable in arousal assessments (Lesage, 2016) and easy to read. With this test, the intensity of valence and arousal perception were rated to assess participants' affective state.

Two HTML page components from lab.js were used for the technical implementation. Figure 3 shows the visualization of these components with the total of ten manikins and ten buttons. While valence is represented in section (A), arousal is represented in section (B). Both scales increase from left to right. The participants were asked to click the button directly below the manakin that was most applicable to them, using the VR controller. According to the study's within-subjects design, the result was five measures of valence and arousal for each participant in total, i.e., one measure before the beginning of the test battery, one after testing the information processing speed in RL, and one after each of the three cognitive performance tests in VR. The means (dependent variables) calculated over the five measures were used for the statistical comparison of the age groups.

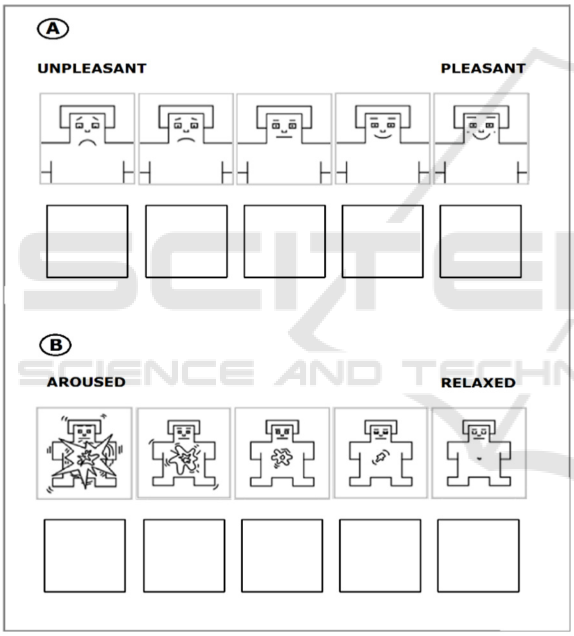

Figure 3: Representation of SAM, showing the selection options for (A) valence and (B) arousal.

2.4 Data Analysis

By using SPSS, the exported data of each measurement phase were verified. First, the data was visualized to reveal any differences. Then, a repeated measures ANOVA (Vincent & Weir, 2012) compared statistically the two consecutive assessments of SAM in detail, using a single-factorial F-distributed function:

$$F(df_1, df_2) = \frac{MSB}{MSW} \tag{1}$$

The given parameters, including the subtraction, correspond to the degrees of freedom. The number of groups, in which the independent variable (age group, affect, gender, and measuring points) can be divided minus one, refers to df_1 (between-groups degrees of freedom estimate). Subtracting one from the number of people in each category and summing across the categories resulted in df_2 (within-groups degrees of freedom estimate). The resulting *F*-value is the ratio of the variance between groups (MSB) to the variance within groups (MSW).

The assumptions needed for using this analysis were met with regard to dependence, scale level of variables, and absence of outliers, but not with regard to normal distribution and sphericity. This function is still intended to check the study's assumptions about the main effects of RL/VR and age by strengthening robustness and reducing any bias with the Greenhouse-Geisser correction (Berkovits et al., 2000).

3 RESULTS

3.1 Descriptive Data

An overview of the participants whose data could be incorporated into the analysis without any technical issues is given in Table 1. The loss of one participant is only due to a single missing response. As several software was connected in series, i.e., VR studio, virtual server machine, and lab.js, the origin of this problem could not be identified. Except this missing response, the data collection was well implemented in terms of software and hardware characteristics.

Table 1: Demographic data of the participants used for the analysis.

	Young	Old
Females	21	15
Males	9	13
Loss	0	1
Total	30	27

3.2 Inferential Statistics

In our 5×2×2 design, two repeated measures analyses of variance were used. In accordance with function (1), the five measuring points for valence and arousal are included as within-subject factors. Age group served as between-subject factor in the first analysis, while this was done with gender in the second analysis.

When looking at the results of the first analysis in a graphical way, a clear difference between the two affective states of valence and arousal can be seen (Figure 4). In both age groups, all five mean scores for arousal were lower than for valence. While the difference is around one score point in the younger group (dark solid line), it is almost two in the elderly (bright solid line). Moreover, the curves show that both groups have a slightly increased stress level after switching to VR. At this point, valence also reaches one of its lowest scores for both groups.

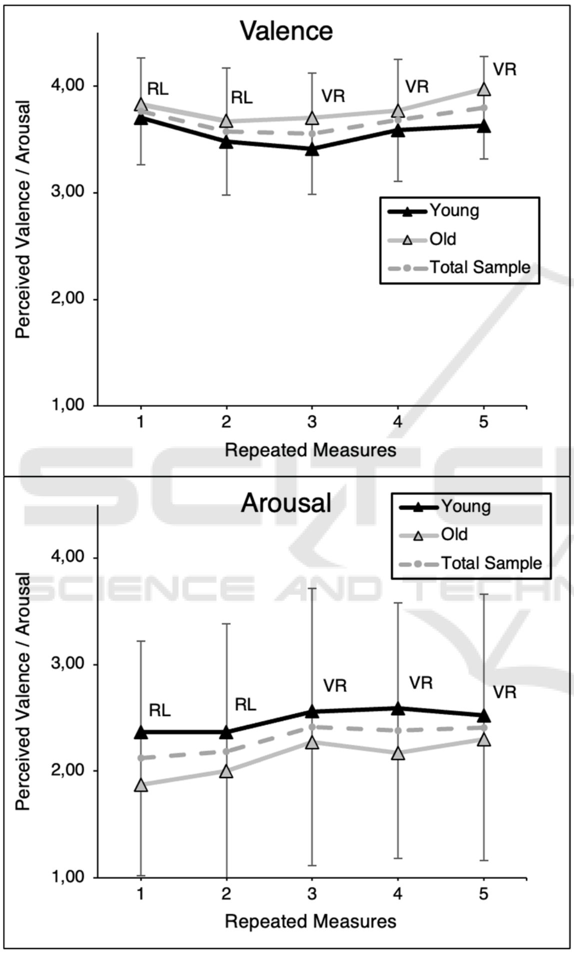

Figure 4: Reported valence and arousal mean scores for young, old and total sample with standard deviation bars for the total sample, in a chronological sequence for the real life (RL) and virtual reality (VR) segments of the study procedure.

Although, there was a clear and statistically significant difference between the affective state scores for valence and arousal within the whole sample, $F(1, 55) = 68.28$, $p < .001$, partial $\eta^2 = .554$, no statistically significant differences could be discovered across the five measurement points, $F(4, 220) = 1.61$, $p = .188$, partial $\eta^2 = .028$. A subdivision into age groups showed similar results, indicating stability in affective state over the test procedure and hence, over switching from RL to a VR assessment. In addition, no significant effect for age, $F(1, 55) = .61$, $p = .439$, partial $\eta^2 = .011$, or gender, $F(1, 55) = .109$, $p = .743$, partial $\eta^2 = .002$, was found, indicating no meaningful differences in valence or arousal between old and young participants or between female and male participants.

4 DISCUSSION

4.1 Key Results

The results of this study showed no statistically significant differences in the affective effects of VR-based test application for cognitive psychology on an age-diverse sample. The expected change in affective state from RL to VR assessment was found neither in the younger nor in the older participants, indicating no influence on the participants' valence or arousal levels when mounting the VR headset and conducting further cognitive tests. Meanwhile, there was no meaningful difference of valence or arousal scores between the young and old age group, indicating a mood stability of this testing environment for different age groups. Overall, it seems that the stressful experience of the participant in VR is identical to that in RL. Alternative explanations might imply the presence of both a higher excitement during the use of a novel technical setup and meanwhile poorer performance results due to distraction effects, that eventually evened out and surely would need separate exploration in future research.

4.2 Limitations

Limitations of this study surely include the explicit, rating-based assessment of the affective state. Possible delusions of the results by high face-validity or interpersonal expectancy effects towards the experimenter (Rosenthal & Rubin, 1978) could not be avoided. Monitoring with the EEG (Electroencephalogram) could improve and stabilize the results.

Furthermore, this study focused on the feasibility of psychological assessments in VR. With regard to this primary step, the small sample size with rather small and imbalanced subgroups in terms of gender and age limits the representative character of our sample and of our findings. The strong overrepresentation of female participants, in particular, could have altered our results, as male individuals often do show a stronger interest in technical innovations and do show lower insecurities towards it (Goswami & Dutta, 2015).

Possible age-related differences could be explored more clearly in future studies if intermediate age groups between 30 and 49 would be included into the sample. In this regard, age should be used as continuous independent variable in the analysis instead of categorizing it into age groups in order to strengthen the result by using an appropriate regression model (Streiner, 2002; Sauerbrei & Royston, 2010).

A systematic bias from participant recruitment cannot be fully ruled out, as the advertisements on digital bulletin boards might have led to a stronger representation of individuals more interested in digital and immersive research technologies.

4.3 Conclusion

This study investigated the effects on participants' cognitive performance and affective state when conducting an assessment in a VR scenario created based on an image of the real world.

Regarding to technical aspects, the lab.js builder showed through a modular principle of basic components its suitability for the development of experimental test in RL as well as in VR if adapted correctly. Even the virtual environment, the handheld controller, or the headset could not disturb the participants. In conclusion, VR environments for cognitive assessment seems to have no significant effect on participants' affective state compared to RL, allowing a promising opportunity for further use of VR without losing any cognitive capacity, e.g., in treating or preventing mental illness.

Other applications of VR in educational contexts have reported contrasting findings (Parong et al., 2021) as the participants affective state indeed showed slightly higher arousal. A possible difference of the presented content and elongation of the performance related context might have resulted in this discrepancy from our findings. However, Holzwarth et al. (2021) reported relatively stable affective states for the entering phase of a VR scenario in their meaningful groundwork of predicting affective states within VR by using the subjects' head movements as predictors.

Admittedly, a more thorough assessment with a larger and more representative sample could produce different results and should therefore be applied in a future study. Nevertheless, the finding is promising as it justifies this new technology to develop new, innovative paradigms for both basic and applied research. In this context, we did already show the feasibility of cognitive performance assessments with an age-diverse sample in a comparable study (Vahle et al., 2021). Other studies were able to successfully induce an avatar age group specific performance difference on physical and cognitive performance domains, e.g., shown by Vahle and Tomasik (2021). Thus, the present research is a crucial groundwork for applying the present and novel technique to the self-reflexive stereotype research, where young participants experience the embodiment of a virtual old age avatar.

ACKNOWLEDGEMENTS

We would like to thank all those who volunteered to participate in our study, the team of *Pointreef* for its continuous technical support of the VR scenario, and Hannah Butt for her support during data collection.

REFERENCES

Anderson-Hanley, C., Arciero, P.J., Brickman, A.M., Nimon, J.P., Okuma, N., Westen, S.C., Merz, M.E., Pence, B.D., Woods, J.A., Kramer, A.F., & Zimmerman, E.A. (2012). Exergaming and older adult cognition: A cluster randomized clinical trial. *American Journal of Preventive Medicine, 42*(2), 109–119. https://doi.org/10.1016/j.amepre.2011.10.016

Berch, D.B., Krikorian, R., & Huha, E.M. (1998). The Corsi block-tapping task: Methodological and theoretical considerations. *Brain and Cognition, 38*(3), 317–338. https://doi.org/10.1006/brcg.1998.1039

Berkovits, I., Hancock, G.R., & Nevitt, J. (2000). Bootstrap resampling approaches for repeated measure designs: Relative robustness to sphericity and normality violations. *Educational and Psychological Measurement, 60*(6), 877–892. https://doi.org/10.1177/00131640021970961

Bradley, M.M., & Lang, P.J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Bridges, D., Pitiot, A., MacAskill, M.R., & Peirce, J.W. (2020). The timing mega-study: comparing a range of experiment generators, both lab-based and online. *PeerJ, 8,* e9414. https://doi.org/10.7717/peerj.9414

Goswami, A., & Dutta, S. (2015). Gender differences in technology usage — A literature review. *Open Journal of Business and Management, 4*(1), 51–59.

Henninger, F., Shevchenko, Y., Mertens, U.K., Kieslich, P.J., & Hilbig, B.E. (2019). *lab.js: A free, open, online study builder.* https://doi.org/10.31234/osf.io/fqr49

Holzwarth, V., Schneider, J., Handali, J., Gisler, J., Hirt, C., Kunz, A., & vom Brocke, J. (2021). Towards estimating affective states in virtual reality based on behavioral data. *Virtual Reality,* 1–14. https://doi.org/10.1007/s10055-021-00518-1

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication, 50*(1), 46–70. https://doi.org/10.1111/j.1460-2466.2000.tb02833.x

Lang, A., Park, B., Sanders-Jackson, A.N., Wilson, B.D., & Wang, Z. (2007). Cognition and emotion in TV message processing: How valence, arousing content, structural complexity, and information density affect the availability of cognitive resources. *Media Psychology, 10*(3), 317–338. https://doi.org/10.1080/15213260701532880

Lesage, F.-X., Berjot, S., Deschamps, F. (2012) Clinical stress assessment using a visual analogue scale. *Occupational Medicine 62*(8), 600–605. https://doi.org/10.1093/occmed/kqs140

Malcolm, B.R., Foxe, J.J., Butler, J.S., & De Sanctis, P. (2015). The aging brain shows less flexible reallocation of cognitive resources during dual-task walking. *Neuroimage, 117*, 230–242. https://doi.org/10.1016/j.neuroimage.2015.05.028

Parong, J., & Mayer, R. E. (2021). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning, 37*(1), 226–241. https://doi.org/10.1111/jcal.12482

Peña, J., Hancock, J.T., & Merola, N.A. (2009). The priming effects of avatars in virtual settings. *Communication Research, 36*(6), 838–856. https://doi.org/10.1177/0093650209346802

Rosenthal, R., & Rubin, D. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences, 1*(3), 377–386. https://doi.org/10.1017/S0140525X00075506.

Safaii, S., Raidl, M., & Beechler, L. (2013). Comparison of a 6 Session virtual world and real world diabetes program. *Journal of Nutrition Education and Behavior, 45*(4), S36. https://doi.org/10.1016/j.jneb.2013.04.097

Sahakian, B.J., & Owen, A.M. (1992). Computerized assessment in neuropsychiatry using CANTAB: Discussion paper. *Journal of the Royal Society of Medicine, 85*(7), 399–402.

Sauerbrei, W., & Royston, P. (2010). Continuous variables: To categorize or to model?. In C., Reading (Ed.). *Data and Context in statistics education: Towards an evidence based society*. Proceedings of the 8th International Conference on Teaching Statistics. Voorburg: International Statistical Institut.

Streiner, D.L. (2002) Breaking up is Hard to Do: The Heartbreak of Dichotomizing Continuous Data. *The Canadian Journal of Psychiatry, 47*(3), 262–266. https://doi.org/10.1177/070674370204700307

Vahle, N.M., & Tomasik, M.J. (2021). *Declines in memory and physical functioning when young adults experience being old in virtual reality*. Manuscript in preparation.

Vahle, N.M., Unger, S., & Tomasik, M.J. (2021). Reaction time-based cognitive assessments in virtual reality – A feasibility study with an age diverse sample. *Studies in Health Technology and Informatics*, *283*, 139–145. https://doi.org/10.3233/SHTI210552

Vickers, D., & Smith, P.L. (1986). The rationale for the inspection time index. *Personality and Individual Differences, 7*(5), 609–623. https://doi.org/10.1016/0191-8869(86)90030-9

Vincent, W.J., & Weir, J.P. (2012). *Statistics in kinesiology* (4th ed.). Champaign, IL: Human Kinetics.