




ENGINEERING
NEGATIVE-RESULT

Skill retention after desktop and head-mounted-display virtual reality training

Alexander Farr^{1,2} , Leon Pietschmann², Paul Zürcher^{2,3} and Thomas Bohné^{2,*}

¹Technical University of Munich, Munich, Germany, ²University of Cambridge, Cambridge, United Kingdom, and ³University of Lincoln, Lincoln, United Kingdom

*Corresponding author. Email: tmb35@cam.ac.uk

(Received 29 October 2022; Revised 05 December 2022; Accepted 08 December 2022)

Abstract

Virtual reality (VR) is increasingly used in learning and can be experienced with a head-mounted display as a 3D immersive version (immersive virtual reality [IVR]) or with a PC (or another computer) as a 2D desktop-based version (desktop virtual reality [DVR]). A research gap is the effect of IVR and DVR on learners' skill retention. To address this gap, we designed an experiment in which learners were trained and tested for the assembly of a procedural industrial task. We found nonsignificant differences in the number of errors, the time to completion, satisfaction, self-efficacy, and motivation. The results support the view that DVR and IVR are similarly useful for learning retention. These insights may help researchers and practitioners to decide which form of VR they should use.

Key words: 3D; skill retention; training; virtual learning; virtual reality

Introduction

Learning with virtual reality (VR) has several benefits, such as improved motivation (Freina & Ott, 2015), experiencing scenarios that are otherwise infeasible, expensive, or dangerous in reality (Scavarelli et al., 2021), and enhanced learning with sensory cues (Psootka, 1995). VR can be distinguished between desktop virtual reality (DVR—using a desktop screen) and immersive virtual reality (IVR—using a head-mounted display). However, it remains largely unclear which type of VR optimally supports learning (Richards & Taylor, 2015; Selzer et al., 2019), particularly as it relates to learning retention. Our experiment focuses on this research gap by comparing the effectiveness of DVR and IVR training on learning retention over time for a procedural skill.

Methods

Hypothesis development

In their meta-analysis, Wu et al. (2020) present that immediate learning was higher for IVR compared to DVR groups. Since initial learning predicts future performance (Arthur et al., 1998; Kamuche & Ledman, 2005), it is expected that:

H1. The number of errors in the physical assembly in the retention test is lower for participants trained with IVR than those trained with DVR.

H2. Time to completion (TTC) of the physical assembly in the retention test is lower for participants trained with IVR than those trained with DVR.

Exceeding goals leads to satisfaction (Kernan & Lord, 1991) and since *H1* and *H2* hypothesize that the IVR group's performance is better, it is predicted:

H3. Participants trained with IVR rate affective learning outcomes higher than those trained with DVR after the retention test as measured by satisfaction.

The same argument applies to self-efficacy as Moores and Chang (2009) found that performance positively correlates with self-efficacy. Both should be higher for IVR:

H4. Participants trained with IVR rate affective learning outcomes higher than those trained with DVR after the retention test as measured by self-efficacy.

Several prior studies report higher motivation when learning with IVR rather than DVR (Makransky et al., 2019; Mouatt et al., 2020). Since Huang et al. (2021) present that motivation from a VR session can be sustained, it is hypothesized that:

H5. Participants trained with IVR rate affective learning outcomes higher than those trained with DVR after the retention test as measured by motivation.

Research approach

To test the hypotheses outlined above, a between-subjects experiment was conducted. A total of 116 participants were recruited, of which $n = 44$ participants successfully completed the entire experiment. The sample in the retention test is predominantly male as the population of students and staff in engineering (the context of our study) is also mostly male (DVR: 92.31%, IVR: 96.97%). The median age for the DVR group was 21 and 19 years for the IVR group. All participants had a normal or corrected-to-normal vision. The groups were also very similar in assembly experience with 30.77% in the DVR and 27.27% in the IVR group reporting having no assembly experience. None of the participants had previously assembled the task used in the experiment.

The research design was similar to the one used by Bohné et al. (2021). Participants were randomly split into two groups and trained in the execution of an unfamiliar industrial assembly task using either DVR or IVR. The research design is shown in [Figure 1](#). The experiment received ethical approval from the University of Cambridge.

While the DVR training was delivered on a standard laptop computer with an external computer mouse, the IVR training was delivered using an Oculus Quest 2 headset. Both groups received the same training in the same VR environment ([Figure 2](#)). The VR environment was an advanced version of the environment used by Bohné et al. (2021). It included a familiarization phase before the main training to account for the pretraining principle (Mayer, 2017).

After the training, participants assembled the physical components in a real-world context. A researcher observed the physical assembly and recorded the performance. Metrics measured included the time to completely assemble the component (TTC) and the number of mistakes made. The latter was determined using a standardized error counting sheet. While participants could ask for a hint at any time, each hint was counted as an error. The assessment procedure was repeated after 14 days (± 2 days) after the initial training to determine the degree of retention for both groups ([Figure 3](#)).

Results

For data cleaning, outliers with a z -score larger than three were excluded (Osborne & Overbay, 2004). The questionnaires included attention check items to ensure participants would not give careless responses. In case of attention check failures, the participant's data were excluded. In case data points

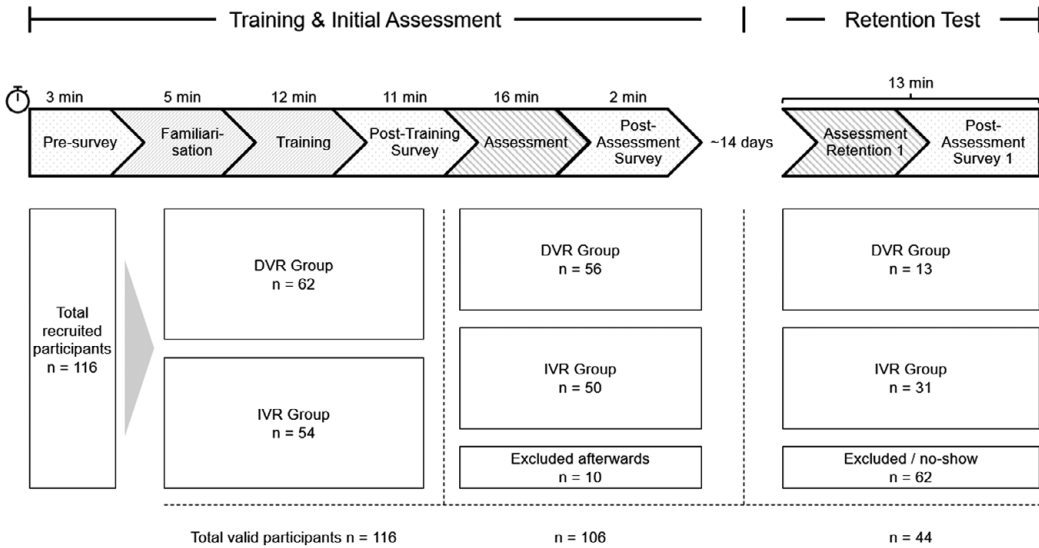


Figure 1. Training procedure and participant flow through the experiment phases.

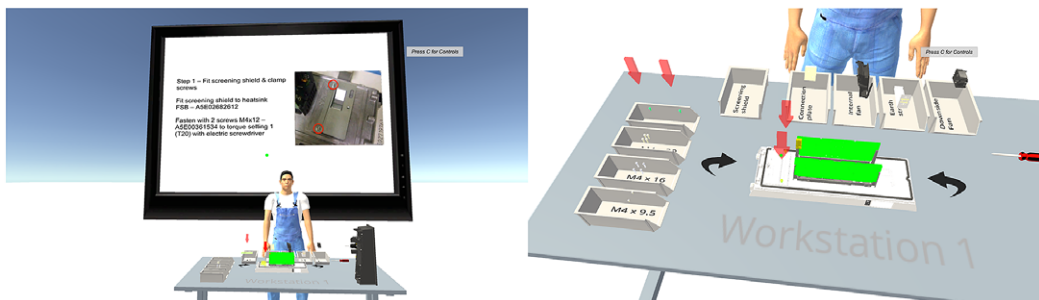


Figure 2. Screenshots from the virtual workstation with instructions on the board behind.



Figure 3. Layout of the physical workstation used for assessment.

for a participant were missing, the data were excluded. Homoscedasticity of data was assessed with Shapiro–Wilk tests and Levene tests were used to assess the normality of the data. If the data were homoscedastic and normally distributed for all groups, ANOVAs were used to compare the data and Wilcoxon rank-sum-tests otherwise. A 5% significance level was used, as recommended by Christensen et al. (2013).

Table 1. Summary of the experiment results

Category	Outcome measure	Hypotheses	Result for assessment in retention test	Means of indicators	
				DVR	IVR
Objective performance factors	Number of mistakes	H1: IVR > DVR	✘ $p = .28$	10.92	10.52
	TTC	H2: IVR > DVR	✘ $p = .47$	637.46	615.81
Affective factors	Satisfaction	H3: IVR > DVR	✘ $p = .42$	5.71	6.14
	Self-efficacy	H4: IVR > DVR	✘ $p = .55$	5.17	5.21
	Motivation	H5: IVR > DVR	✘ $p = .20$	5.65	6.06

Note. A cross indicates that the associated hypothesis is rejected. The number of mistakes was counted manually during the physical assembly. TTC is in seconds. Satisfaction, self-efficacy, and motivation were each assessed with multiple 7-point Likert questions. Abbreviations: DVR, desktop virtual reality; IVR, immersive virtual reality; TTC, time to completion.

Table 1 summarizes the results of the hypotheses tests and the descriptive indicators. All tests are nonsignificant. All descriptive indicators indicate a better performance of the IVR group as the number of mistakes and TTC are lower and the affective factors have been rated higher by the IVR group.

Discussion

The hypotheses that the IVR group would perform significantly better in the retention test than the DVR group as measured by the number of mistakes and TTC were based on a meta-analysis by Wu et al. (2020). They reported a small effect size for better performance of IVR than DVR groups directly after training. However, the studies comparing learning retention after DVR and IVR training are inconclusive as Buttussi and Chittaro (2018) and Lu et al. (2022) found nonsignificant differences, Smith et al. (2018) reported mixed results, and Alrehaili and Al Osman (2019) recorded DVR as slightly better media. Since there are nonsignificant differences in the objective factors for learning retention, the goal attainment should be similar between both groups. The nonsignificant difference in satisfaction in retention is thus expected (Kernan & Lord, 1991). The groups' nonsignificant differences in self-efficacy are in line with Moores and Chang's (2009) finding as there are nonsignificant differences in prior performance. The hypothesis that motivation would be higher in the retention test was more exploratory since little evidence on the relationship has been published so far.

Our experimental results can help researchers to further guide their research. Our study differs from previous and related works because we included retention and directly compared IVR and DVR environments. Contrary to prior studies, participants in our experiment were also assessed by using a physical assembly and thus tested for their procedural and conceptual knowledge rather than only tested for their conceptual knowledge with simple online tests. Furthermore, we ensured that the DVR and IVR training environments were as similar as possible, only differing in their human-computer interaction method.

A limitation of our study is the relatively small sample size in the retention test. Only 44 participants participated in the retention test compared to 106 valid data points for the initial assessment. This discrepancy was the result of the Covid-19 pandemic in the UK at the time of the experiment as many participants had to quarantine during the retention test or were not available for the additional assessment. Additionally, as the population in the training center was predominantly male, the same applies to our sample (3 women and 103 men in the initial test). Results might be different for a more balanced sample.

Three further areas of research are proposed. First, the future experiment should replicate according to the original plan with several retention tests and sufficient participants in the retention tests, as this has still been rarely analyzed in the literature. Second, the transferability of the results should be investigated

by running the experiment with a different training content to test if the findings hold outside an industrial training context. Third, retention of VR training should be compared with retention after learning with other media such as paper or videos.

Conclusion

The results of our experiment suggest that there are nonsignificant differences in performance in retention after learning with DVR or IVR and that these technologies could be used interchangeably. This applies both when measuring performance with performance indicators such as TTC and number of errors, and with affective indicators such as satisfaction, self-efficacy, and motivation. Because of the Covid pandemic, the sample in our study was smaller than planned which might explain the nonsignificant results. Decisions for a media (DVR or IVR) should consider the results of this study and other factors that were not examined in our study including cost, software development complexity, the appropriateness of a task for DVR or IVR, and the required infrastructure for DVR or IVR training.

Acknowledgments. The authors would like to thank Helen Dudfield, Mayowa Olonilua, and Thomas Govan for their advice in planning the experiment, and Athena Hughes, James Tombling, Julius Bock, Kim Ruf, Konrad Ciezarek, Siemens Congleton, and MakeUK for supporting the experiment.

Open peer review. To view the open peer review materials for this article, please visit <http://doi.org/10.1017/exp.2022.28>.

Data availability statement. The data that support the findings of this study are available upon request from the corresponding author.

Authorship contributions. A.F., L.P., P.Z., and T.B. conceived and designed the study. P.Z. adapted the Virtual Reality application to the requirements of the experiment. A.F. and L.P. conducted the data gathering. A.F. and P.Z. analyzed the data. A.F., L.P., P.Z., and T.B. wrote the article.

Funding statement. This work was supported by the Defence Science and Technology Laboratory (Dstl) under its Serapis framework. Permission to Publish was granted by Dstl in May 2022.

Conflict of interest. The authors have no conflicts of interest to declare.

References

- Alrehaili, EA & AIOsman, H (2019): A virtual reality role-playing serious game for experiential learning. In *Interactive Learning Environments*, pp. 1–14. DOI: [10.1080/10494820.2019.1703008](https://doi.org/10.1080/10494820.2019.1703008).
- Arthur, W., Jr., Bennett, W., Jr., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance*, *11*, 57–101. https://doi.org/10.1207/s15327043hup1101_3
- Bohné, T., Heine, I., Gurerk, O., Rieger, C., Kemmer, L., & Cao, L. Y. (2021). Perception engineering learning with virtual reality. *IEEE Transactions on Learning Technologies*, *14*, 500–514. <https://doi.org/10.1109/TLT.2021.3107407>
- Buttussi, F & Chittaro, L (2018): Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario. In *IEEE Transactions on Visualization and Computer Graphics* *24* (2), pp. 1063–1076. DOI: [10.1109/TVCG.2017.2653117](https://doi.org/10.1109/TVCG.2017.2653117).
- Christensen, L. B., Johnson, B. R., & Turner, L. A. (2013). *Research methods, design, and analysis* (12th ed.). Pearson, pp. 433–441
- Freina, L., & Ott, M. (2015). A literature review on immersive virtual reality in education: State of the art and perspectives. In *Proceedings of the 11th international scientific conference eLearning and software for education* (Vol. *133*). <https://www.semanticscholar.org/paper/A-LITERATURE-REVIEW-ON-IMMERSIVE-VIRTUAL-REALITY-IN-Ott-Freina/e93b38f3892c7357051f39be6b6574f298a3b72a>.
- Huang, W., Roscoe, R. D., Johnson-Glenberg, M. C., & Craig, S. D. (2021). Motivation, engagement, and performance across multiple virtual reality sessions and levels of immersion. *Journal of Computer Assisted Learning*, *37*, 745–758. <https://doi.org/10.1111/jcal.12520>
- Kamuche, F. U., & Ledman, R. E. (2005). Relationship of time and learning retention. *Journal of College Teaching & Learning*, *2*, 25–28. <https://doi.org/10.19030/tlc.v2i8.1851>
- Kernan, M. C., & Lord, R. G. (1991). An application of control theory to understanding the relationship between performance and satisfaction. *Human Performance*, *4*, 173–185. https://doi.org/10.1207/s15327043hup0403_1

- Lu, F., Nanjappan, V., Parsons, P., Yu, L., & Liang, H.-N.** (2022). Effect of display platforms on spatial knowledge acquisition and engagement: An evaluation with 3D geometry visualizations. *Journal of Visualization*, 1–20. <https://doi.org/10.1007/s12650-022-00889-w>
- Makransky, G., Borre-Gude, S., & Mayer, R. E.** (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning*, 35, 691–707. <https://doi.org/10.1111/jcal.12375>
- Mayer, R. E.** (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, 33, 403–423. <https://doi.org/10.1111/jcal.12197>
- Moore, T. T., & Chang, J. C.-J.** (2009). Self-efficacy, overconfidence, and the negative effect on subsequent performance: A field study. *Information & Management*, 46, 69–76. <https://doi.org/10.1016/j.im.2008.11.006>
- Mouatt, B., Smith, A. E., Mellow, M. L., Parfitt, G., Smith, R. T., & Stanton, T. R.** (2020). The use of virtual reality to influence motivation, affect, enjoyment, and engagement during exercise: A scoping review. *Frontiers in Virtual Reality*, 1, 564664. <https://doi.org/10.3389/frvir.2020.564664>
- Osborne, J. W., & Overbay, A.** (2004). The power of outliers (and why researchers should ALWAYS check for them). <https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1139&context=pape>
- Potka, J.** (1995). Immersive training systems: Virtual reality and education and training. *Instructional Science*, 23, 405–431. <https://doi.org/10.1007/BF00896880>
- Richards, D., & Taylor, M.** (2015). A comparison of learning gains when using a 2D simulation tool versus a 3D virtual world: An experiment to find the right representation involving the Marginal Value Theorem. *Computers & Education*, 86, 157–171. <https://doi.org/10.1016/j.compedu.2015.03.009>
- Scavarelli, A., Arya, A., & Teather, R. J.** (2021). Virtual reality and augmented reality in social learning spaces: A literature review. *Virtual Reality*, 25, 257–277. <https://doi.org/10.1007/s10055-020-00444-8>
- Selzer, M. N., Gazcon, N. F., & Larrea, M. L.** (2019). Effects of virtual presence and learning outcome using low-end virtual reality systems. *Displays*, 59, 9–15. <https://doi.org/10.1016/j.displa.2019.04.002>
- Wu, B., Yu, X., & Gu, X.** (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology*, 51, 1991–2005. <https://doi.org/10.1111/bjet.13023>

Peer Reviews


Reviewing editor: Dr. Guney Guven Yapici

Ozyegin University, Mechanical Engineering, Nisantepo Mh. Orman Sk. No:34-36, Istanbul, Turkey, 34794

Minor revisions requested.

doi:10.1017/exp.2022.28.pr1

Review 1: Skill retention after Desktop and Head-Mounted-Display Virtual Reality training

Reviewer: Dr. Hai-Ning Liang 

Date of review: 21 November 2022

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Conflict of interest statement. Reviewer declares none.

Comment

Comments to the Author: In this paper, the authors present their work in comparing 3D immersive HMD against 2D desktop-based display for learning retention of an assembly task. Overall, the paper is well-written and presents the work in a succinct manner. The paper also cites relevant papers from the literature. Below is another paper that could be used in the discussion section, as it is closely related to their work.

Lu et al. Effect of display platforms on spatial knowledge acquisition and engagement: an evaluation with 3D geometry visualizations. *J Vis* (2022). <https://doi.org/10.1007/s12650-022-00889-w>.

Another limitation to be pointed out is the male-dominated population, in addition to the small sample size due to pandemic restrictions. As such, it is unclear if the same effects would be found if there is a more balanced population ratio.

Also, I wonder if the authors interviewed (even briefly) the participants after the experiment. If they did, it would be useful to add some of this analysis.

Score Card

Presentation



Is the article written in clear and proper English? (30%)

5/5

Is the data presented in the most useful manner? (40%)

4/5

Does the paper cite relevant and related articles appropriately? (30%)

4/5

Context



Does the title suitably represent the article? (25%)

5/5

Does the abstract correctly embody the content of the article? (25%)

5/5

Does the introduction give appropriate context? (25%)

4/5

Is the objective of the experiment clearly defined? (25%)

5/5

Analysis



Does the discussion adequately interpret the results presented? (40%)

5/5

Is the conclusion consistent with the results and discussion? (40%)

5/5

Are the limitations of the experiment as well as the contributions of the experiment clearly outlined? (20%)

4/5

Review 2: Skill retention after Desktop and Head-Mounted-Display Virtual Reality training

Reviewer: Dr. Nigel Newbutt 

University of Florida, Gainesville, Florida, United States, 32611-7011

Date of review: 28 November 2022

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Conflict of interest statement. Reviewer declares none.

Comment

Comments to the Author: This is a sound, and timely, research paper. The details are clearly outlined and the experiment well designed and articulated. The foreground is a little limited in its length and the context for this study could be more clearly drawn out. But the findings are insightful. I also would have liked to see confirmation of ethical review and procedures therein. I found the focus on the ‘significance’ to be a little limiting too — after all the IVR experience was better in many of the measures taken (just not significantly). This could have been reflected in the limitations and also findings section. But the results and conclusions are inline with the hypotheses of this particular experiment. I’d like to congratulate the authors for a great contribution to the field, and for working with a large sample during the pandemic. Thanks for the chance to preview this important work.

Score Card

Presentation



Is the article written in clear and proper English? (30%) 5/5

Is the data presented in the most useful manner? (40%) 4/5

Does the paper cite relevant and related articles appropriately? (30%) 4/5

Context



Does the title suitably represent the article? (25%) 5/5

Does the abstract correctly embody the content of the article? (25%) 5/5

Does the introduction give appropriate context? (25%) 3/5

Is the objective of the experiment clearly defined? (25%) 5/5

Analysis



Does the discussion adequately interpret the results presented? (40%) 4/5

Is the conclusion consistent with the results and discussion? (40%) 4/5

Are the limitations of the experiment as well as the contributions of the experiment clearly outlined? (20%) 4/5