Classification of the Topicality and Relevance of Evaluation Tools for VR Applications

Caroline Schon [1], Rong Huang[2], Hanna Hessenmüller[1], Simon Przybyl[1] and Johannes Tümler[1]
[1]*Department Electrical Engineering, Mechanical Engineering and Engineering Management, Anhalt University of Applied Sciences, Bernburger Str. 55, 06366 Köthen, Germany*
[2]*Department Architecture, Facility Management and Geoinformation, Anhalt University of Applied Sciences, Franz-Mehring Str. 21, 06846 Dessau, Germany*
{*caroline.schon, rong.huang, hanna.hessenmueller, johannes.tuemler*}*@hs-anhalt.de, simon.przybyl@student.hs-anhalt.de*

Keywords: Virtual Reality, Evaluation, Questionnaire Analysis, User Involvement

Abstract: Virtual reality (VR) has emerged as one of the fastest growing technologies in recent years, offering transformative opportunities in various disciplines such as education, medicine and industry. Classifying the advantages and disadvantages of such VR simulations must be done using appropriate evaluation tools. Despite these rapid technological developments, the evaluation of VR experiences lags behind technological advances. This article examines the discrepancy between the rapid technical evolution of VR and the slow progress in the development of suitable evaluation methods. Existing evaluation approaches are analyzed and their limitations in the context of new VR applications are identified. In addition, potential solutions and future research directions will be identified to enable a more appropriate assessment of the user experience, effectiveness and impact of VR technologies. The aim is to promote an integrated evaluation framework that can keep pace with technological innovations and thus support the sustainable use of VR in different application areas.

1 INTRODUCTION

With the promise of offering experiences that immerse users in simulated worlds, virtual reality (VR) has created applications that were previously unimaginable. The potential applications of VR are diverse and range from education and healthcare to architecture and entertainment. This opens up new horizons for innovation and interaction. The technical development of VR is taking place at a rapid pace. Advances in graphics, sensors and computing have led to significant improvements in the performance and accessibility of VR systems. Current VR headsets offer very good image quality and interactive options that were considered unattainable just a few years ago. These technological innovations have promoted the acceptance and integration of VR in various industries and defined new standards for immersive experiences.

Effective evaluation methods are essential to optimize the user experience, evaluate effectiveness in specific applications and identify potential risks or undesirable effects. Although the technical possibilities are developing at a rapid pace, the evaluation of VR experiences is still challenging. The majority of existing evaluation methods were originally designed for the evaluation of traditional media formats and therefore suggests that they are unable to comprehensively capture the specific characteristics and potential of VR. The need to develop adequate evaluation methods that take into account the complexity and multilayered nature of VR is becoming increasingly apparent.

The praxwerk project[1] of Anhalt University of Applied Sciences focuses on the digitalization of higher education, with the integration of VR into teaching and learning processes as a central research focus. As part of the project, the virtual learning application VR-BioTech-House® has been further developed to present complex biotechnological processes in a practical and interactive manner. This bridges theoretical knowledge with practical application, thereby enhancing the achievement of learning objectives. However, a key challenge is to systematically evaluate the impact of VR, particularly in terms of its effectiveness and its contribution to specific learning objectives. Evidence-based analysis is essential to fully unlock the potential of VR in higher education and guide future developments.

The aim of this paper is to provide both re-

[1]https://hs-anhalt.de/praxwerk

searchers and practitioners working on similar applications such as the BioTech-House, with an overview of the current state of evaluation in VR research and to provide impetus for future developments.

2 LITERATURE

In a meta analysis 14 different articles were reviewed that dealt with the evaluation of mixed reality (xR) and especially VR applications in a wide variety of areas with a focus on applications with a purpose of education.

Springer, Google Scolar and the Discovery Service of the Library of Anhalt University of Applied Sciences were used as search platforms for the analysis with variations of the search terms: “evaluation XR application”, “evaluation VR application” and “quality assessment VR application”. The articles found through this method were checked to see whether they dealt sufficiently with the explicit phrasing of different categories relevant to the evaluation. First and foremost, the research was based on articles that questioned the evaluation itself, but in order to obtain a larger number of studies, additional articles were included that evaluated specific applications as long as they explicitly mentioned the criteria used for evaluation.

Consequently, seven overarching categories were identified from the examined articles, which will be discussed in more detail below. These are:

- user experience;
- user assistance;
- game (or application) mechanics;
- quality factors;
- graphics;
- VRise (VR sickness or VR induced symptoms and effects);
- motivation.

Each of these categories contains smaller factors, although the number of these varies depending on the category and not each of them has been mentioned in every source. However, it should be noted that it is difficult to clearly delineate these categories, as the quality of the graphics, for instance, can have a direct influence on the user experience.

Factors that can be categorized as user experience were mentioned in a total of ten articles [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] whereby the extent to which this category was subdivided into smaller factors varies greatly depending on the source material. For example, user experience is generally listed twice as a factor [5, 6]. Overall, the following factors were found, that have an influence on the factor user experience: an adequate level of immersion [1], pleasant VR experience [1], high quality sounds [1]/ audio [2], suitable hardware (head mounted display (HMD) and computer) [1], information gift [2] (the way information is presented), language [2], ease to use [2, 3, 7], similarity to reality [3] and factors that depend on the user themselves: cognitive skills - e.g., visuospatial abilities etc., [4], participants attitude [4], levels of trust/acceptance of VR/MR tools [4, 9], motivation during usage [4, 7, 9], previous experience with VR (XR) [4, 5, 8] and levels of presence and engagement [4, 6, 10].

The greatest importance is attributed to ease of use, the user factor levels of presence and engagement, previous experience and motivation during usage, which are each mentioned in three different sources, as well as audio quality and the user factor levels of trust/acceptance of VR/MR tools with two mentions each. All other factors were only found in one source each.

The second identified category is user assistance. Factors belonging to this category were mentioned in three sources [1, 2, 6] and describe in which way the user is introduced to or guided through the program. These include the following factors: digestible tutorials [1], helpful tutorials [1], adequate duration of tutorials [1], (helpful in-game) instructions [1, 2], helpful in game prompts [1], goal and task design [2], feedback [2], information transmission [2] and didactics of in-game instructions [6].

As a result, (helpful in-game) instructions are assigned the greatest importance. This factor is mentioned directly in two of the three sources, although the factor of the last source relevant here, didactics of in-game instructions, also intersects this factor.

The next category, game mechanics, includes factors from six sources that determine movement and interaction in the application. These are the following: a suitable navigation system (e.g. teleportation) [1] / locomotion [2] / navigation operation [11], availability of physical movement [1], naturalistic picking/ placing of items [1], naturalistic use of items [1], naturalistic 2-handed interaction [1], interaction elements [2], user interface [2, 7], operation of interaction functionalities [2], design/functionalities of interaction functionalities [2], relevance assessment of interaction functionalities [2], menu control [2], control systems [2], room design [2], tracking [2], haptic interaction [3] and interaction regulation [12].

The factor most frequently acknowledged is the way in which the navigation system is implemented, having been mentioned in three different sources. The factors user interface and haptic interface were also referred to repeatedly with two instances each.

The quality factors of a VR application largely include factors that are difficult to measure in units and are often rather subjective. The basic factors referred to are dependability, validity, objectivity, reliability, effectivity, knowledge, utility and VR technology, all of which are discussed by Napetschnig [2]. Utility and knowledge are also dealt with by Tsivitanidou [6] and the latter by Antonopoulos[5]. Other factors are system performance [3, 8], system reliability [3] and technical aspects and tools features - e.g., effect of designed features, expected and experienced system functioning [2, 6].

The greatest importance is therefore assigned to the knowledge factor, i.e. the quality in which the application conveys knowledge to the user, which was discussed by Chardonnet [11]. This is followed in order of importance by utility, VR technology, system performance and the technical aspects and tools features, each of which are mentioned by Yoon [13].

The quality of the graphical representation [2] includes all aspects of an application that relate to the graphical representation. It includes the sub-aspects of software, degree of accuracy, structure, fineness/detail, concept, design, texture, value, claim and 3D character or character appearance. These aspects can be found in Napetschnig [2], whereby degree of accuracy can also be assigned to Kojić [7] and 3D character or character appearance, design and texture to Wienrich [8]. Finally, the aspects of the 2D model display and the 3D display, which are assigned to Ye [3], should also be mentioned.

The greatest importance was placed on the aspects of degree of accuracy [2, 7], design [2, 8], texture [2, 8] and 3D character/avatar appearance [2, 8]. In general, it can be seen that realism and accuracy of the representation play a major role among users. It is important to mention that the aspects of graphical representation dealt with were not always clearly definable, particularly in the case of Wienrich [8], as some of the evaluating participants made inaccurate descriptions of their experience with the simulation. An attempt was made to classify this information as well as possible in the given categories.

VRise (VR sickness or VR induced symptoms and effects) is considered here under the aspect of negative physiological reactions to VR applications. In the course of this, all of the following aspects of Kourtesis [1] are dealt with, as well as absence or insignificant presence of nausea, absence or insignificant presence of disorientation and cybersickness by Wang [14]. The latter is the only part of this list to be found in Chardonnet [11] and the absence or insignificant presence of dizziness in Yoon [13]. The remaining parameters are absence or insignificant presence of fatigue and absence or insignificant presence of instability.

The greatest importance can obviously be attributed to cybersickness, which includes many of the other sub-aspects. It is therefore mentioned directly in three sources, while the absence or insignificant presence of nausea [1, 14], absence or insignificant presence of disorientation [1, 14] and absence or insignificant presence of dizziness [1, 13] are only mentioned in two sources each.

Finally, motivation in use will be discussed. It describes the aspects of a VR application that motivate its use. Napetschnig [2], Borsci [4], Wienrich [8] and Kojić [7] address these aspects. Napetschnig [2] also refers to the subcategories logging, configuration, adaptability/customizability for different needs and interests, which is also mentioned by Kojić [7], practicability [8, 9], user-friendliness [7, 8] and usability [3, 8]. Learnability is only mentioned by Ye [3].

Practicability [2, 8, 9], user-friendliness [2, 7, 8] and usability [2, 3, 8] were mentioned the most of these factors, apart from the factor of motivation in use in general, which means that they are assigned the greatest importance.

It is obvious that the aspects for evaluating VR applications are not only numerous, but also cover a spectrum of sub-areas. Not only are issues as the visual and acoustic representation of the immersive world important, but also the effect of this world on the user, which in turn can be broken down into different aspects such as physiological reactions, the learning effect or sensory impressions. However, the question arises as to whether this selection of criteria is sufficient for the rapidly advancing state of VR technology or whether the questionnaires are already outdated compared to the demand for the evaluation of VR applications.

3 EVALUATION METHOD

The question posed will be answered using a qualitative research approach. This is particularly suitable for the present study, as it has a small sample size and an explorative character. The latter aims to gain deeper insights into user perceptions.

The data was collected as part of an experiment conducted during a science camp with pupils at the age of 13-15 years. Two groups of twelve technology and science enthusiasts took part in the study. These participants represent the target group that will increasingly come into contact with VR technologies

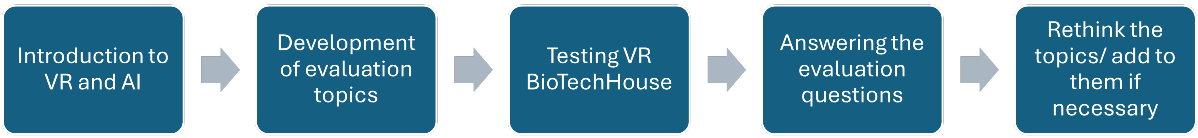

Figure 1: Procedure of the survey.

in the future. The application VR-BioTech-House was tested by the young people in an educational application with the goal of developing categories for evaluating the application. The implementation was part of a workshop on the topic of xR and AI. Firstly the students were introduced to xR and AI, then they were tasked with the development of evaluation criteria with the help of ChatGPT in groups of two. Only after that they had the oportunity to test the VR application BioTech-House which is presented below. Finally after the exploration phase, the criteria of all groups were discussed and if necessary amended or more added as can be seen in Figure 1.

The VR-BioTech-House application is an educational tool in the field of biotechnology that allows learners to explore everyday applications of biotechnology within a virtual simulation of an ordinary house. Participants can engage in hands-on activities to produce biotechnology products, such as yogurt [15], red wine, and yeast for bread-making [16]. In this experiment, the yogurt-making scenario was selected as the test case.

The central instruments of this study were the VR scenario, which served as the object of the experiment, and the online notepad, which the student groups used to record their criteria. The evaluation criteria thus found are the data of this study. The duration of the testing phase of the VR-application for each person was 25 minutes. During that, they could explore the application freely and did not get any outside instructions unless they asked for it. The goal was a successful completion of the scenario.

4 RESULTS

The results of the qualitative analysis revealed diverse user needs and expectations for VR applications, focusing on hardware, software, and other contributing factors. These insights emphasize the importance of aligning the design and evaluation of VR systems with these requirements to enhance user satisfaction and functionality.

The first thematic block focuses on hardware, as illustrated in Figure 2. An important point here is robustness, which refers to the hardware's resistance to physical stress. The sensor accuracy or tracking is also relevant concerning the precision with which the sensors detect and reproduce movements. Another aspect to consider is sustainability and fair trade, which aim to ensure environmentally friendly manufacturing and fair production conditions. The connections, in particular connectivity, are also relevant. The latter includes the type and number of connections as well as possible wireless connections. The field of vision provides information about the area that users can see through the hardware. The term comfort covers both the ergonomics and the wearing comfort of the devices. Another topic is the controllers. In this context, functionality is of central importance, whereby the operation and input options of the hardware are of crucial relevance. Finally, the sensory impressions are also relevant, which are influenced by audio and graphics, especially color, and determine the sensory perception of the user.

In the second thematic block, the results on the topic of software are sumarized and can be seen in Figure 3. A key aspect is user-friendliness, which refers to the ease of use and intuitive operability of the software. The software's ability to support multiple users simultaneously is addressed in the multiplayer category. The ability to customize the setting and content to meet the specific requirements and needs of the user is also of particular relevance. Adaptation to the respective learning level and consideration of different learning types are particularly important here. Another aspect to consider in this context is the performance of the software. It is of interest how smoothly it runs and how stable it is in use. The integration of visual or auditory effects and surprises can have a positive influence on the user experience. The ability of users to actively interact with the software is another important aspect that is subsumed under the term interactivity. The quality of the content offered by the software is recorded under the Content category. In this context, the fun of using the software and the topicality of the content, which is ensured by regular updates, are also emphasized. Furthermore, the focus on a specific target group is relevant, as adapting the software to the needs of this user group is essential.

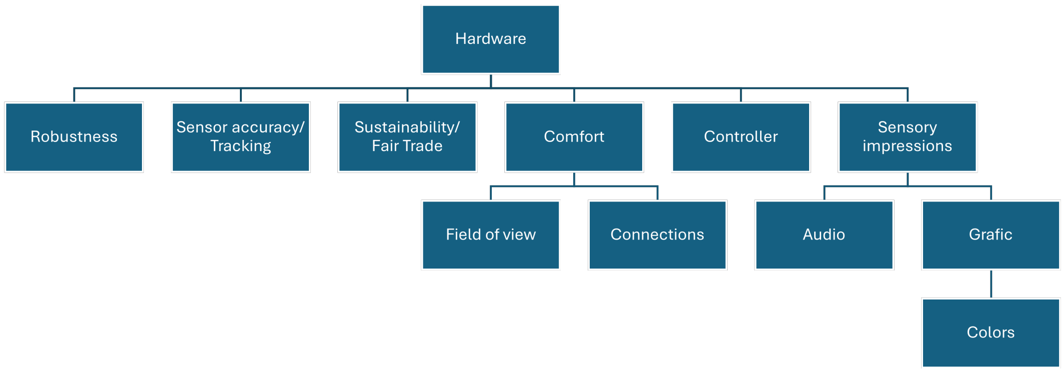

Figure 2: Results with a focus on hardware.

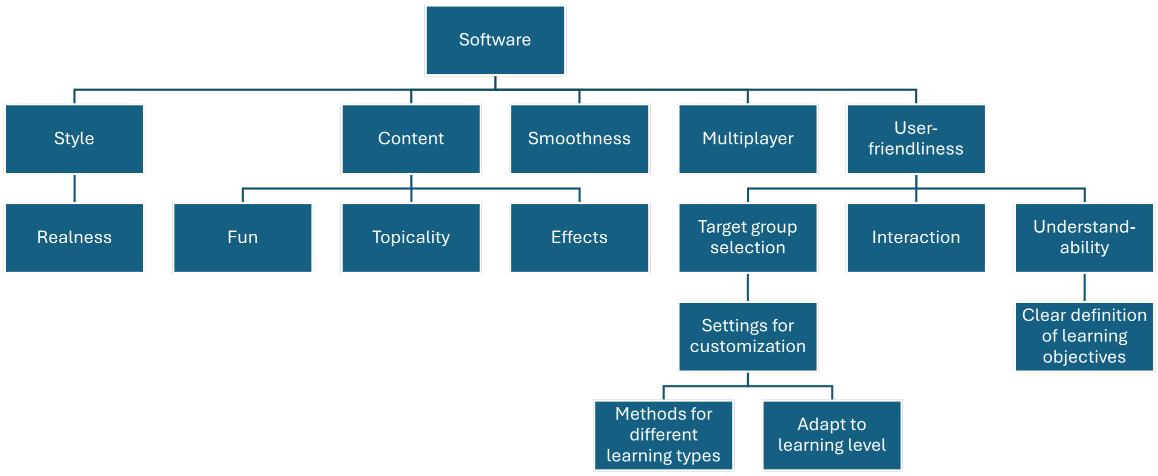

Figure 3: Results with a focus on software.

The aspect of comprehensibility of the application includes the adequate explanation of the functions and content of the software. In this context, the precise definition of the learning objectives is of crucial importance. Finally, the style of the software is also relevant, as is the extent to which the application is designed realistically.

In the last block, all topics that relate to neither hardware nor software are assigned. These are summarized in Figure 4. In the context of the use of VR, the aspect of health is also relevant, whereby the effects of the use of VR on health must be examined. It is also necessary to determine how efficient the use is, in particular the suitability of the experience in the context of learning and its potential to promote long-term engagement. Another aspect is the feeling or gaming experience evoked by the app, as well as the difficulty of use and integration into everyday teaching. The next aspect discussed was accessibility, with a particular focus on adjustable language options.

After testing, most participants did not add any new categories. However, a few added topics such as task and learning management as well as language and room boundaries. One group wrote: “We would definitely use other criteria now, e.g. content, audio”.

Compared with the findings of the literature review, it is notable that most criteria found during the review were mentioned by the different student groups in one way or another. The depths of the description may have varied, but the general implication was present. Notable for these similarities are for example user-friendliness as found in the evaluation with the criteria ease of use and intuitive operability of the software, with the corresponding criteria that appeared during the literature review in the categories user-experience (ease of use) and Motivation in use (user-friendliness). Notable differences between the literature review and the BioTech-House evaluation are that different areas were emphasized. During the evaluation, a clear category of hardware was formulated while during the literature review the criteria that are influenced by hardware are sub-aspects of bigger categories. Furthermore, the evaluation put more emphasis on a game-characteristic (Feel-

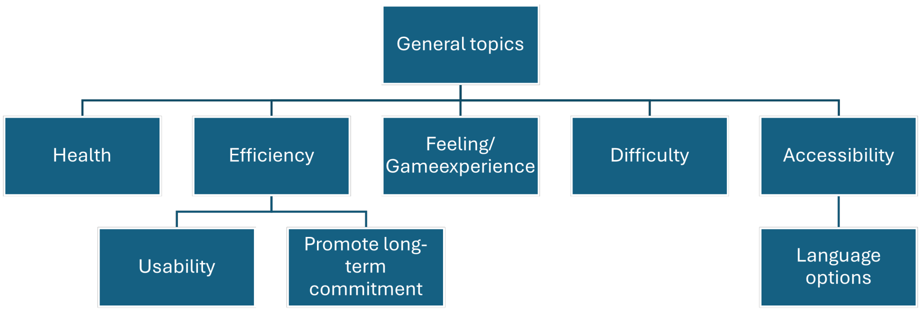

Figure 4: Results with a focus on general topics.

ing/Gameexperience, Fun), while the closest criteria of the literature review is motivation in use. The only two categories unique to the evaluation are the criteria multi-player and sustainability.

The findings provide key insights into the future development of BioTech-House. Hardware considerations, such as robustness, comfort, and precise tracking, are essential for creating a reliable and immersive learning experience. On the software side, user-friendly interfaces, adaptive learning features, and high-quality content are crucial to making virtual experiments both engaging and pedagogically effective. The participants also stressed the importance of interactivity, real-time feedback, and accessibility to ensure inclusivity and usability in diverse educational contexts.

In general, these findings offer practical guidance for the evaluation and improvement of BioTech-House. By addressing user needs in both hardware and software design, the application can optimize its effectiveness as a virtual learning tool. Furthermore, the results contribute to a broader understanding of how to develop VR systems that are both engaging and educationally impactful.

5 DISCUSSION

This analysis provides a comprehensive summary of user needs and expectations regarding the BioTech-House educational application, offering clear directions to optimize hardware, software, and overall user experience. In particular, the discussion of hardware emphasizes the importance of comfort, graphical and audio presentation, and robustness. These attributes are indispensable for tools used in educational settings, especially by students, where durability and reliability under varying conditions are critical.

Insights on software usability provide actionable guidance for developers working to enhance the application. Key factors include user-friendliness, support for multiplayer interactions, and the ability to customize educational content. Additionally, efforts to improve content engagement and maintain high-quality learning materials are essential to meet the needs of university students and other learners. Addressing aspects such as tailored content, adaptive difficulty levels, and seamless interactivity is crucial to ensure the software's educational effectiveness.

The results of this qualitative analysis align closely with the practical requirements of the BioTech-House project. These findings offer a roadmap for refining the user experience and strategically integrating features desired by users, thereby supporting the next stages of development. By addressing these considerations, BioTech-House can better fulfill its goal of delivering an engaging and effective virtual learning tool.

6 CONCLUSIONS

This study investigates the relevance of existing evaluation categories of VR experiences. It highlights the need for diverse evaluation parameters that encompass immersion, spatial perception, and psychological reactions. The literature review suggests that existing methods take these aspects into account, but the results indicate that this study may be limited by biases due to AI systems or participant influence. Furthermore, due to the specific background of the participating students as people with an interest in science and technology it is possible that they did not think of difficulties people with a lesser interest would have or put a greater emphasis on the abilities of hardware.

To improve the evaluation, the development of an integrative framework combining methods from different disciplines such as psychology, computer science and design is proposed. Interdisciplinary approaches seem particularly suitable for comprehensively mapping the complex dimensions of VR experiences. Future research directions could investigate how machine learning can be used to personalize VR experiences and adapt the evaluation in real time, they could explore if a greater focus on the impact of the used hardware in the evaluation could be helpful.

A sustainable and adaptable development of these evaluation strategies is seen as crucial to realize the full potential of VR technologies. Close collaboration among survey designers, developers, and end-users is essential to bridge the gap between technological advancements and user feedback.

ACKNOWLEDGMENTS

We thank the German "Stiftung Innovation in der Hochschullehre" for partially funding this work.

REFERENCES

[1] P. Kourtesis, S. Collina, L. A. A. Doumas, and S. E. MacPherson, "Validation of the virtual reality neuroscience questionnaire: Maximum duration of immersive virtual reality sessions without the presence of pertinent adverse symptomatology," Frontiers in Human Neuroscience, vol. 13, no. 417, pp. 1-15, Nov 2019, [Online]. Available: https://doi.org/10.3389/fnhum.2019.00417.

[2] A. Napetschnig, K. Brixius, and W. Deiters, "Development of a core set of quality criteria for virtual reality applications designed for older adults: Multistep qualitative study," Interactive Journal of Medical Research, vol. 12, pp. 1-15, Sep 2023, [Online]. Available: https://doi.org/10.2196/45433.

[3] J. Ye, S. Badiyani, V. Raja, and T. Schlegel, "Applications of Virtual Reality in Product Design Evaluation," in Human-Computer Interaction. HCI Applications and Services, J. A. Jacko, Ed., pp. 1190-1199, 2007, [Online]. Available: https://doi.org/10.1007/978-3-540-73111-5_130.

[4] S. Borsci, G. Lawson, and S. Broome, "Empirical evidence, evaluation criteria and challenges for the effectiveness of virtual and mixed reality tools for training operators of car service maintenance," Computers in Industry, vol. 67, pp. 17-26, Feb 2015, [Online]. Available: https://doi.org/10.1016/j.compind.2014.12.002.

[5] P. Antonopoulos, E. Fokides, and G. Koutromanos, "Understanding learning and learning experience in immersive virtual reality," Tech Know Learn, pp. 1-20, Aug 2024, [Online]. Available: https://doi.org/10.1007/s10758-024-09764-z.

[6] O. E. Tsivitanidou, Y. Georgiou, and A. Ioannou, "A learning experience in inquiry-based physics with immersive virtual reality: Student perceptions and an interaction effect between conceptual gains and attitudinal profiles," Journal of Science Education and Technology, vol. 30, no. 6, pp. 841-861, Dec 2021, [Online]. Available: https://doi.org/10.1007/s10956-021-09924-1.

[7] T. Kojic, D. Ali, R. Greinacher, S. Möller, and J.-N. Voigt-Antons, "User experience of reading in virtual reality - finding values for text distance, size and contrast," in 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), pp. 1-6, 2020, [Online]. Available: https://doi.org/10.1109/QoMEX48832.2020.9123091.

[8] C. Wienrich, N. Döllinger, S. Kock, K. Schindler, and O. Traupe, "Assessing User Experience in Virtual Reality - A Comparison of Different Measurements," in Design, User Experience, and Usability: Theory and Practice, A. Marcus and W. Wang, Eds., pp. 573-589, 2018, [Online]. Available: https://doi.org/10.1007/978-3-319-91797-9_41.

[9] N. Pares and R. Pares, "Towards a model for a virtual reality experience: The virtual subjectiveness," Presence: Teleoperators and Virtual Environments, vol. 15, no. 5, pp. 524-538, Oct 2006, [Online]. Available: https://doi.org/10.1162/pres.15.5.524.

[10] P. Grabarczyk and M. Pokropski, "Perception of affordances and experience of presence in virtual reality," AVANT. The Journal of the Philosophical-Interdisciplinary Vanguard, vol. 7, no. 2, pp. 25-44, Dec 2016, [Online]. Available: https://doi.org/10.26913/70202016.0112.0002.

[11] M. A. Mirzaei, F. Merienne, and J.-R. Chardonnet, "Influence of navigation parameters on cybersickness in virtual reality," Virtual Reality, vol. 25, no. 3, pp. 565-574, Sep 2021, [Online]. Available: https://doi.org/10.1007/s10055-020-00474-2.

[12] A. Halbig and M. E. Latoschik, "A systematic review of physiological measurements, factors, methods, and applications in virtual reality," Frontiers in Virtual Reality, vol. 2, pp. 1-20, Jul 2021, [Online]. Available: https://doi.org/10.3389/frvir.2021.694567.

[13] H. J. Yoon, J. Kim, S. W. Park, and H. Heo, "Influence of virtual reality on visual parameters: immersive versus non-immersive mode," BMC Ophthalmology, vol. 20, no. 1, pp. 1-10, Dec 2020, [Online]. Available: https://doi.org/10.1186/s12886-020-01471-4.

[14] G. Wang and A. Suh, "User adaptation to cybersickness in virtual reality: A qualitative study," in Proceedings of the 27th European Conference on Information Systems (ECIS), pp. 1-12, Jun 2019, [Online]. Available: https://aisel.aisnet.org/ecis2019_rp/191.

[15] R. Huang, J. Rödigh, A. Berger, and J. Tümler, "Ein virtuelles biotech-haus für das lehren in der biotechnologie," in Proceedings of DELFI Workshops 2022, pp. 77-85, 2022, [Online]. Available: https://doi.org/10.18420/delfi2022-ws-17.

[16] J. Tümler, M. Alptekin, K. Krylova, R. Huang, and J. Rödigh, "Weiterentwicklung des vr-biotech-house: Wein- und hefeherstellung," in Proceedings of DELFI Workshops 2024, pp. 1-10, 2024, [Online]. Available: https://doi.org/10.18420/delfi2024-ws-04.