Towards Effective Virtual Reality Learning Environments: Assessment of Information Processing and Learning through Eye Tracking

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
M.Sc. Philipp Stark
aus Esslingen am Neckar

Tübingen
2024

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	01.07.2024
Dekan:	Prof. Dr. Thilo Stehle
Berichterstatter/-in:	Prof. Dr. Enkelejda Kasneci
Berichterstatter/-in:	Prof. Dr. Michiko Sakaki

"Commit to an idea that is bigger than yourself"
- inspired by Simon Sinek.

To my wife, my family, and my friends

Acknowledgments

First of all, I would like to thank my supervisors, Prof. Dr. Enkelejda Kasneci and Prof. Dr. Richard Göllner, for the great opportunity to do my dissertation in this interdisciplinary field and for all the support during my doctoral studies. Thank you very much for your help, advice, guidance, and mentoring during my doctorate. I wish to extend my sincerest appreciation to Prof. Dr. Michiko Sakaki for her willingness and effort in evaluating my thesis. Further, special thanks and greatest appreciation go to Prof. Dr. Michael Krone and Dr. Shahram Eivazi for their immediate willingness to evaluate my work and my research.

Furthermore, I would like to express my heartfelt gratitude to the following individuals and institutions who have contributed significantly to the completion of this dissertation:

I am sincerely thankful to Dr. Efe Bozkir and Dr. Hong Gao for their expertise, support, and guidance on all matters related to eye-tracking and VR. Your collaboration has been invaluable to the success of this study. I would like to express my sincere appreciation to Dr. Wolfgang Fuhl, Dr. Thomas Kübler, and Dr. Tobias Appel for their prompt assistance, valuable insights, and guidance on technical and research-related matters. Your willingness to support my scientific endeavors has been truly invaluable. At this point, again, a very special thanks to Prof. Dr. Enekelejda Kasneci for her great guidance and provision of knowledge and resources crucial to my development.

Further, I would like to extend my sincere gratitude to Prof. Dr. Ulrich Trautwein and the Hector Research Institute of Education Sciences and Psychology for providing me with the opportunity to pursue my dissertation. I deeply appreciate the stimulating work environment and abundant opportunities for personal and professional growth during my time at the institute. At this point, again, a very special thanks to Prof. Dr. Richard Göllner for the extensive support, insightful discussions on my research, and for facilitating connections with individuals who offered invaluable assistance with my inquiries.

Furthermore, I would like to express my sincere appreciation to Dr. Lisa Hasenbein and Dr. Joseph Ferdinand for their contribution to my projects and for generously sharing data. Your efforts in experiment design and data collection, as well as your support, have made my

research possible. I would like to thank Babette Bühler, Hannah Deininger, and Alexander Jung for their unwavering emotional support as office mates and for patiently addressing numerous of my questions throughout this journey. I am deeply grateful to Dr. Gabe Orona, Dr. Alexandr Ten, Julia-Kim Walther, and Wy Ming Lin for the enriching discussions, fruitful exchange of ideas, and overall support each of you provided.

I would like to thank Prof. Dr. Jens-Uwe Hahn for the invaluable technical support with VR, assistance in conceptualizing ideas, and enabling the creation of VR environments essential to this research. My heartfelt thanks also go to Prof. Dr. Wolfgang Schnotz for the stimulating discussions and invaluable feedback on our research design.

Additionally, I would also like to extend my gratitude to the LEAD Graduate School and Research Network, particularly to Dr. Simone Oechslen, Sophie Freitag, Elena Sizana, and Alisa Schmid for their invaluable support in organizational, procedural, and administrative matters. Your contributions and the vibrant exchange within the research network have been instrumental in my growth and development.

Finally, I would like to express my infinite gratitude to my wife, Marlies, who has been incredibly supportive during this doctoral period. I would like to thank my parents, Monika and Wolfgang, for their years of support and their great interest in my scientific journey. I would also like to thank Moritz, Felix, Jan, and my brother Dominik for their advice, feedback, and emotional support during my doctorate. And thanks to many others who have accompanied me during this time and helped to enrich my work.

Philipp Stark

Abstract

Due to recent technological innovations, virtual reality (VR) has become a promising technology for systematically investigating learning in more authentic scenarios while simultaneously providing a controllable experimental setting. When bridging the gap between standardized lab experiments and real-life phenomena, eye tracking can be a rich source of information. Eye tracking can be instrumental for assessing information processing and learning in VR, and analyzing visual attention through eye tracking can provide valuable insights for creating effective virtual learning environments.

However, investigating eye tracking in 3D environments also poses some challenges, including integrating head movement, acquiring gaze target information, and interpreting eye movements in relation to information processing and learning. This thesis addresses some of these challenges by proposing methodological and analytical solutions, including reliable measures of pupil diameter, gaze-ray casting, network analysis, gaze entropy, and machine-learning models. These approaches are used to measure information processing and learning through eye-tracking and explore the potential for modeling eye movements and visual attention.

First, two standardized virtual experiments focus on processing and encoding information with 3D objects and measuring reliable pupil diameter baselines in VR. Second, the visual attention distribution in a virtual classroom is analyzed to understand the effects of the classroom environment and different teaching events on the students. In the last step, gaze-based attention networks are utilized to study the effect of social-related behavior on visual attention and learning in VR classrooms.

This work contributes to expanding knowledge of VR research in education science and explores the possibilities of eye-tracking analysis in VR. The findings aim to offer insights into information processing and learning in virtual environments and contribute to developing effective virtual learning environments.

Zusammenfassung

Dank neuesten technologischen Innovationen sind virtuelle Realitäten (VR) zu einer vielversprechenden Technologie geworden, um das Lernen in authentischeren Szenarien systematisch zu untersuchen und gleichzeitig eine kontrollierbare Versuchsumgebung zu schaffen. Um die Lücke zwischen standardisierten Laborexperimenten und realen Phänomenen zu schließen, kann Eye Tracking eine reichhaltige Informationsquelle sein. Die Analyse der visuellen Aufmerksamkeit durch Eye Tracking kann wertvolle Erkenntnisse für die Konzeption effektiver virtueller Lernumgebungen liefern.

Die Analyse von Eye Tracking in 3D-Umgebungen birgt jedoch auch einige Herausforderungen, darunter die Integration von Kopfbewegungen, die Erfassung von Gaze-Target Informationen und die Interpretation von Augenbewegungen in Bezug auf Informationsverarbeitung und Lernen. Diese Dissertation befasst sich mit einigen dieser Herausforderungen und schlägt methodische und analytische Lösungen vor, die eine zuverlässige Messung des Pupillendurchmessers, Gaze-Ray Casting, Netzwerkanalyse, Gaze Entropy und maschinelle Lernmodelle umfassen. Diese Ansätze werden zur Messung der Informationsverarbeitung und des Lernens durch Eye-Tracking genutzt und zeigen ein Potenzial für die Modellierung von Augenbewegungen und visueller Aufmerksamkeit.

Zuerst konzentrieren sich zwei standardisierte virtuelle Experimente auf die Verarbeitung und Kodierung von Informationen mit 3D-Objekten und die Messung reliabler Baselines für Pupillendurchmesser in VR. Zweitens wird die Verteilung der visuellen Aufmerksamkeit im virtuellen Klassenzimmer analysiert, um die Auswirkungen des Klassenraums und verschiedener Lehrszenarien auf die Schülerinnen und Schüler zu verstehen. Im letzten Schritt werden Gaze-based Attention Networks eingesetzt, um den Effekt von sozialem Verhalten auf die visuelle Aufmerksamkeit und das Lernen im virtuellen Klassenzimmer zu untersuchen.

Diese Arbeit trägt zur Erweiterung des Wissens von VR-Forschung in der Bildungsforschung bei und erforscht die Möglichkeiten der Eye-Tracking-Analyse in VR. Die Ergebnisse sollen Einblicke in die Informationsverarbeitung und das Lernen in virtuellen Umgebungen bieten und zur Entwicklung effektiver virtueller Lernumgebungen beitragen.

Contents

List of Figures

List of Tables

Acronyms

AOI Area of Interest. 11

FOV Field of View. 10

GBDT Gradient Boosting Decision Tree. 16

HMD Head-Mounted Display. 4

I-DT Dispersion Identification Threshold. 16

I-VT Velocity Identification Threshold. 16

OOI Object of Interest. 11

SHAP Shapley Additive Explanations. 17

VR Virtual Reality. 3

1. List of Publications

Accepted Publications Relevant to this Thesis

[1] **P. Stark**, E. Bozkir, W. Sójka, M. Huff, E. Kasneci, and R. Göllner, "The impact of presentation modes on mental rotation processing: A comparative analysis of eye movements and performance", *Scientific Reports*, 2024. DOI: 10.1038/s41598-024-60370-6

[2] **P. Stark**, A. Tobias, O. Milo, and K. Enkelejda, "Pupil diameter during counting tasks as potential baseline for virtual reality experiments", in *2023 Symposium on Eye Tracking Research and Applications (ETRA '23)*, Germany: ACM, Jun. 30, 2023, p. 7. DOI: 10.1145/3588015.3588414

[3] E. Bozkir, **P. Stark**, H. Gao, L. Hasenbein, J.-U. Hahn, E. Kasneci, and R. Göllner, "Exploiting object-of-interest information to understand attention in VR classrooms", in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, Mar. 2021, pp. 597–605. DOI: 10.1109/VR50410.2021.00085

[4] **P. Stark**, A. Jung, J.-U. Hahn, E. Kasneci, and R. Göllner, "Using gaze transition entropy to detect classroom discourse in a virtual reality classroom", in *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications (ETRA '24)*, Glasgow, UK: ACM, 2024. DOI: 10.1145/3649902.3653335

[5] **P. Stark**, L. Hasenbein, E. Kasneci, and R. Göllner, "Gaze-based attention network analysis in a virtual reality classroom", *MethodsX*, vol. 12, p. 102 662, Jun. 1, 2024. DOI: 10.1016/j.mex.2024.102662

[6] L. Hasenbein, **P. Stark**, U. Trautwein, A. C. M. Queiroz, J. Bailenson, J.-U. Hahn, and R. Göllner, "Learning with simulated virtual classmates: Effects of social-related configurations on students' visual attention and learning experiences in an immersive virtual reality classroom", *Computers in Human Behavior*, vol. 133, p. 107 282, Aug. 1, 2022. DOI: 10.1016/j.chb.2022.107282

Other Publications not Relevant to this Thesis

[7] J. Ferdinand, H. Gao, **P. Stark**, E. Bozkir, J.-U. Hahn, E. Kasneci, and R. Göllner, "The impact of a usefulness intervention on students' learning achievement in a virtual biology lesson: An eye-tracking-based approach", *Learning and Instruction*, vol. 90, p. 101 867, Apr. 1, 2024. DOI: 10.1016/j.learninstruc.2023.101867

[8] H. Gao, E. Bozkir, **P. Stark**, P. Goldberg, G. Meixner, E. Kasneci, and R. Göllner, "Detecting teacher expertise in an immersive VR classroom: Leveraging fused sensor data with explainable machine learning models", in *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Oct. 2023, pp. 683–692. DOI: 10.1109/ISMAR59233.2023.00083

[9] L. Hasenbein, **P. Stark**, T. Ulrich, H. Gao, E. Kasneci, and R. Göllner, "Investigating social comparison behaviour in an immersive virtual reality classroom based on eye-movement data", *Scientific Reports*, vol. 13, no. 1, p. 14 672, Sep. 6, 2023. DOI: 10.1038/s41598-023-41704-2

2. Introduction

Due to recent technological innovations, Virtual Reality (VR) celebrated a rebirth in the consumer market with immersive, head-mounted VR devices at affordable prices [10]. The elevated level of immersion allows individuals to detach from their actual surroundings and fully submerge into the virtual environment [10]. Consequently, VR is becoming an increasingly relevant technology for educational practice [11] and enables the investigation of scientific questions in education science and psychology [12]–[15]. It unlocks the possibility of investigating participants' behavior in more authentic scenarios than provided by conventional lab experiments, and it bridges the gap between results from standardized lab experiments and real-life phenomena in the world that researchers are keen to understand. More precisely, VR enhances ecological validity while concurrently providing a standardized and controlled experimental setting [16]–[18]. This unique combination also raises the interest of VR for education science [19], [20] and opens up the question of the effective utilization of VR learning environments for research and practice [21]–[28].

Effectiveness in virtual learning environments can be viewed from different perspectives. VR can be an effective training environment that simulates a real-life learning situation [29]. It can be an effective learning environment in which students learn specific concepts and skills [30] that are more difficult to acquire in other learning environments [31]–[33]. However, effectiveness can also be understood in terms of intuitive access to learning environments in which students demonstrate authentic behavior without additional or less effort [1], [4], [7], [34]–[37]. At the same time, an effective learning environment can also be one in which learning behavior can be analyzed and monitored efficiently [5], [38]–[40] and which is particularly suitable for psychological testing [2], [41]–[45].

To approach the topic of the effectiveness of virtual learning environments, a focus should be placed on the essential functions of VR, which create a unique virtual experience. The VR experience encompasses aspects of perception, learning, and social cognition, for which specific technical developments form the foundation. First, VR creates the perception of a 3D space by simulating binocular disparity and motion parallax as two prominent depth

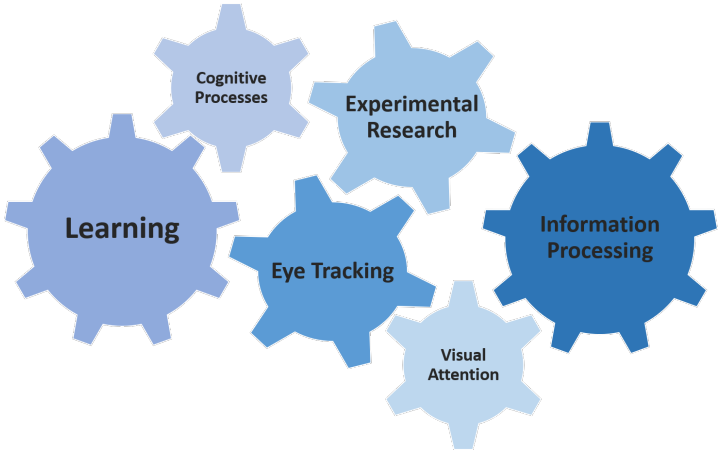


Figure 1.: Connected concepts and framework of this thesis

cues [46], [47]. Second, recent advances in animation and scene rendering allow us to create dynamic VR environments that give the user a feeling of immersion and presence [48]. Because these environments are made with 3D game engines, users can freely move around and explore the scene, which gives them a feeling of space [49]–[51]. Developments like motion capture enable the creation of a lively virtual environment with virtual characters that evoke a sense of social presence [52], [53].

These possibilities of designing an immersive VR environment simultaneously lead to a trade-off between experimental rigor and ecological validity [54]. A greater degree of freedom for the participants in a VR experiment can be assumed to lead to more diverse individual behavior. To investigate the behavioral differences of the participants, further measurement methods must be used to collect data in the process. This process data can describe, predict, and explain human behavior in virtual environments and allow researchers to draw conclusions about cognitive processes [55]. In this context, eye tracking can be a rich source of information for investigating information processing and learning. Eye tracking in VR is a non-invasive assessment method [56] that can provide valuable information without disrupting the virtual experience. Eye-tracking data can be used to capture and analyze participants' visual attention, which allows for an objective assessment of their activities and abilities [57] and provides insight into cognitive processes associated with the learning experience [58].

While many affordable VR Head-Mounted Display (HMD) on the consumer market come with integrated eye trackers, valuable analysis of eye-tracking data in VR environments is

challenging [59], [60]. The limited spatial accuracy, precision, and low temporal resolution do not allow for the analysis of all types of eye movement and area of interest information. Further, free head movement imposes a challenge when detecting fixations, while motion and lightning in the scene require additional methods to obtain reliable measures of gaze targets and pupil diameter. Since visual attention is also highly context- and environment-dependent [61], [62], previous findings on interpreting patterns and distributions of visual attention need further evaluation in new virtual settings. At the same time, these challenges make it possible to break new ground in eye-tracking analysis and explore new methods for modeling visual attention in virtual environments.

In the following sections, a more detailed motivation for the scientific contributions of this thesis is provided. Section 2.1 elaborates on using VR in education science and introduces specific learning environments like the virtual classroom. It describes the technical and theoretical foundations of the virtual experience to understand what constitutes learning in VR. Section 2.2 focuses on the theoretical foundation of information processing and learning, motivating the importance of visual attention for learning and introducing the link to eye-tracking analysis. Given that VR imposes some challenges regarding analyzing eye-tracking data, Section 2.3 explains the basic concepts of VR eye-tracking and motivates the methods used to model eye movements and visual attention in this thesis. A visualization of the connected concepts and a framework thesis is depicted in Figure 1.

2.1. Virtual Reality Learning Environments

In recent decades, more and more studies have used VR in psychology and education science [26]. VR has been extensively utilized to study teacher education and training [21], [30], learning across all age groups from preschool to higher education [27], [63], domain-specific learning [20], [23], [64], learning impairments [12] or pedagogy [28]. Thereby, different virtual learning environments were employed, with the virtual classroom being a prominent example [43], [65]–[67].

VR depends on some technical basics that affect our perception when using these devices. These technical basics need to be considered if we want to understand the challenges of analyzing eye tracking in VR and formulate implications for information processing and learning. The term virtual reality typically refers to the technology where users wear an HMD headset, in which images are projected in real-time on stereoscopic displays [68]. An HMD occludes all visual information from the real world and enables users to submerge

into a virtual world [19], [69], which is referred to as immersive VR [32]. In contrast to mixed or augmented reality, VR usually refers to creating an artificial environment that provides no real-time combination of the physical and the virtual world [70]. Sometimes VR is also defined as the virtual environment itself [69], where users can experience a reality simulation that is perceived as "almost real" [12]. In this thesis, the term VR always refers to an immersive virtual reality experience with an HMD.

Two main technical basics are utilized to create the perception of depth in the virtual space. First, the stereoscopic displays in an HMD generate a perception of stereoscopic depth [68]. This is produced by a slight displaying offset that aligns with the binocular disparity of the eyes [47] and contributes to identifying spatial relations in 3D spaces. Second, the position and rotation of the HMD in space are tracked. Users receive feedback on their head movements by constantly updating the displays to reflect the user's expected motion parallax in 3D spaces [10], [47]. The possibility of perspective and location changes create sensory-motoric and spatial immersion [71]. It enables users to perceive the three-dimensional structure of objects and provides them with a sense of place [47]. This is particularly important, as the perception and manipulation of 3D objects improve their recognition and recall [72]. However, there is still a lack of understanding of the effects of stimulus depth on visual behavior and experience in virtual environments [73].

Additionally, immersive VR allows for free locomotion. Walking around freely in VR is only possible if the environment is designed and created as a 3D environment. In contrast to 360° videos [74], which only show a video, VR simulations are created by game engines [16], [75], which allow for 3D rendering of a virtual space. This gives the user the feeling of being placed in a true-to-perspective 3D environment in which places and objects have a spatial position and relation to each other [75]. This makes VR a valuable tool to study and assess cognitive processes [76] and visuospatial abilities during the perception of 3D objects [77].

The two key concepts describing a VR virtual experience are immersion and presence [12]. The concept of immersion can be understood in different ways. From a technical perspective, immersion refers to an "inclusive, extensive, surrounding and vivid illusion of reality" [78] (p.3). From the users' perspective, immersion can also be defined as a psychological state characterized by perceiving oneself to be surrounded by and interacting with the virtual environment [78]. The latter definition also allows different levels of immersion to be distinguished. The more a person is immersed in the virtual world, the higher the perceived level of immersion [48].

Presence in a virtual environment can be understood as the subjective experience and

illusion of being there in a virtual place, referred to as spatial presence [49], [79]. Presence is a key element affecting participants' cognition and learning in VR [80]. Realism is another aspect of presence that describes the experience of plausibility for events occurring in VR. This is closely linked to social presence, which considers the plausibility of social interaction in VR and gives the users the feeling of the social presence of another person [81]. The social information processing theory [82] highlights the manifold ways of social behavior in virtual learning spaces, which also includes social interaction with animated None-Player Characters [49], [83]. Because we are social animals [84], small social cues of animated characters can already generate social responses, even though we are aware that the characters are not real [49]. The combination of behavioral realism and perceived agency of the virtual social agents can influence individuals' social comparison [9], [39], social understanding and conformity [85], [86], as well as performance in the presence of virtual social characters [87]–[89]. Using VR to study social-related behavior is especially appealing because researchers can retain experimental control while investigating complex social situations [52].

Although VR allows us to immerse ourselves in other worlds [24], exploring everyday activities can contribute to understanding learning and social comparison [39]. The classroom, for example, is the central learning environment for students, which contributes to their emotional, cognitive, and academic development through social interaction and social relationships between its participants [90]. For this reason, the virtual classroom is a frequently investigated virtual environment [26], [65]. It serves as an environment to study students' attention [91]–[93], attention disorder [76], [94] and cognitive performance [67], [95]–[97]. Classroom design aspects [37], [43], [66], [98], classroom climate [99], teaching [64] and teacher expertise [8] are further aspects that have been investigated in virtual classrooms.

The different studies show that VR provides high ecological validity to assess learning in various situations [26], [100]. However, VR research is also inconclusive when it comes to the effectiveness of learning in VR [32], [74], [101], [102]. Some studies treat VR as a black box, with learning in VR only assessed after the experiment through questionnaires. To investigate the learning behavior of the participants during the VR experience, physiological process measures should be traced [55]. Analyzing participants' eye movements and visual attention might be the most straightforward approach, given that some VR devices come with integrated eye trackers, and analyzing eye-tracking data can reveal information about participants' information processing and learning [62].

2.2. Assessment of Information Processing and Learning

Since learning is a fundamental part of being human, many different perspectives, concepts, and theories exist. In this thesis, we look at learning from the perspective of information processing [103]. As part of the cognitive learning theories, information processing focuses on the internal mental processes and how humans encode, store, and retrieve information during learning. Therefore, perception, memory, and knowledge acquisition are integral to this theory. Through the lenses of information processing, the idea of internal representations helps explain how we process information. Representations refer to internal cognitive mechanisms that represent all our knowledge about the world inside our minds[103], [104]. Our activities and abilities emerge as the interaction of internal representations and the world around us [105]. Although this includes the processing of all sensory information, visual information processing plays a particularly important role.

Since vision is our primary sense, humans' eyes are a rich source of information when studying information processing [106]. Visual attention can provide insight into information encoding, visual strategies, or social comparison processes. Certain paradigms guide the research on visual attention.

First, visual attention can be defined as a selective process that describes allocating limited attentional resources to specific information in the visual field while ignoring other information [107]. In the concept of a spotlight, visual attention acts as a gatekeeper for visual working memory [108], [109]. This selective mechanism is necessary due to the limited capacity of processing visual information [110] and the competition with other sensory information [111]. Thus, attention and learning have a codependent relationship whereby attention acts as a selective mechanism that facilitates the learning process [112]. This connection between visual attention and knowledge acquisition also becomes evident when looking at the anatomy of the eye. Due to the foveal system of the eye, the highest visual acuity is limited to a small central area of the retina [113], [114].

Second, visual attention can manifest as overt when it aligns with an individual's eye movement toward a specific location. This means that an individual's focus of attention coincides with their eye fixation [56]. Lab experiments showed that recall and memory are better for longer fixated objects during scene perception, further establishing the link between visual attention and knowledge acquisition [115]. According to the eye-mind hypothesis, there is temporal alignment between what is fixated and what is processed in the brain. Lab experiments have shown that the duration of fixations reflects perceptual intake and processing

[116], [117].

Third, visual attention can be separated into two categories, top-down and bottom-up visual attention [118]. While bottom-up visual attention is guided by salient visual stimuli appearing in the surrounding [118], experiments like the Yarbus task [119] have strongly emphasized the role of top-down attention. Different fixation patterns for different task instructions on the same stimuli indicated that humans show voluntary eye movements towards locations important for them [119], [120]. This leads to the assumption that humans have partial control of their own learning and can process information meaningfully.

Last, another aspect of information processing is cognitive load [121]. The cognitive load theory assumes that the capacity of the working memory is limited and strained differently during learning. Among other measurements, the contraction and dilation of the pupil serve as an indicator of cognitive load or arousal, which can affect the working memory and therefore learning in educational environments [122]–[128]. Generally, a higher cognitive load is associated with an increased pupil diameter [115].

Eye tracking is a non-invasive physiological measurement method that can capture human eye movements to measure visual attention [115]. The utility of this method is that their raw variables (e.g., gaze direction) can be extrapolated to constructs associated with information processing and learning [75]. Eye movement features can be calculated to investigate cognitive processes during reading [116], [129], (spatial) problem-solving [130]–[133] or scene perception [129], [134], [135]. Besides the study of individual eye movement features, visual scanning patterns and gaze transition information can be used to study individual differences in various tasks [136]–[143] as well as the joint attention of multiple individuals [144]. Further, gaze-based networks can be applied to study individual differences in visual perception [145]–[147] and collaboration behavior [148]–[150].

As a consequence, information acquired by eye-tracking is broadly applied in fields like cognitive science [75], psychology [135], [151] and human-computer interaction [152]. In the field of education science, visual attention measured by eye tracking is used in manifold ways to study attention and learning [62], [130], [153]–[157]. Eye tracking also indicates social components of visual attention [84], [158] and social interactions [144], [159]. Further, eye tracking is employed to obtain a reliable measurement of cognitive load to investigate aspects of task difficulty, the design of learning materials, or the assessment of individual competencies [35], [57], [112], [160]–[167].

Given the variety of eye-tracking applications for investigating information processing and learning, it is promising to apply this method in virtual realities. However, there are some

challenges when analyzing eye-tracking data in VR.

2.3. Eye Tracking in Virtual Reality

Implementation and application of eye tracking in VR are considered to be different from other eye-tracking methods [75]. Presumably, due to the relatively recent development of integrated eye-tracking systems in VR headsets, there are still few standard software solutions available to obtain processed eye-tracking data and calculated eye movement features [73]. This requires researchers to develop their own solutions for data collection and processing. Before tapping into the challenges of processing and analyzing eye-tracking data in VR, the technical basics of eye-tracking in VR should be described.

The HMD used in all studies of this thesis was the HTC VIVE Pro Eye [168], where each of the binocular displays provided a resolution of $1440x1600$ pixels per eye with a 110° Field of View (FOV) and a refresh rate of 90 Hertz (Hz). All data was collected via the integrated Tobii eye tracker. Although this setup represented state-of-the-art technology in the field of VR eye tracking, there were also some limitations in terms of technical capabilities. According to company specifications, this eye tracker has a trackable FOV of 110°, a self-reported accuracy of 0.5°- 1.1° within the 20° FOV [168], and a 5-point eye-tracking calibration. However, studies about the accuracy and precision of this eye tracker reported a lower accuracy, especially in the periphery of the visual field, and an influence of head movement on the precision of the eye tracker [169], [170]. Further, this eye tracker provided a frame rate of $120Hz$, which corresponds to one data point roughly every 8.3 millisecond. In comparison, remote eye trackers like the EyeLink1000 [171] provide a frame rate of $2000Hz$. However, when combining the eye-tracking data with information gathered in the VR environment (e.g., head movement), the temporal resolution is bound to the frame rate of the VR environment. Because the frame rate of the VR environment depends on the complexity of the rendered scene, the movement of the participants, and the performance of the computer, this can reduce the frame rate to even below $50Hz$ (one data point every 20 millisecond) [58].

The accuracy, precision, and temporal resolution of VR eye trackers limit the possibility of eye-tracking analysis and do not allow the calculation of low amplitude eye movements such as microsaccades [58] or saccade duration [172]. However, some eye movement features have been applied in VR eye-tracking experiments [58]. For example, pupil diameter has been measured to indicate cognitive load in different virtual learning scenarios [173]. Fixation- and saccade-related features have been analyzed in VR to measure attention, cognitive state, and

emotional response [174]. The free head movement of participants adds another layer of complexity when analyzing eye tracking in VR. Algorithms for fixation detection need adjustments to incorporate head movement [58], [175]. This means that the meaningful interpretation of eye tracking in VR is always a combination of eye and head-related features. Furthermore, to account for the non-linear relationship between these features, machine learning algorithms and explainability approaches can provide additional support in interpreting the results [176]–[178].

Another important aspect is that the gaze direction in the virtual environment is calculated as a combination of the local gaze direction recorded by the integrated eye tracker and the head position and rotation of the HMD in the virtual space. The gaze target location can then be obtained by using gaze-ray casting [179]. While there are already proposed software solutions implemented within Unity, there is no such solution for the Unreal Engine [75]. The gaze-ray casting method allows one to obtain information about the Object of Interest (OOI), which is similar to an Area of Interest (AOI), the object the gaze is targeting. A ray that represents the participant's gaze is cast into the virtual environment and collides with a 3D object in the environment, the so-called gaze target. With this method, one can analyze the gaze duration for specific OOIs and the gaze transitions between the objects. Such information has, for example, been studied to indicate visual attention during learning tasks to distinguish expertise levels [180]. One advantage of analyzing gaze transitions in VR is that one doesn't have to calculate fixations. Gaze transitions between objects and OOI duration can be directly obtained by the gaze ray-casting methods. For example, transition information between AOIs was used to calculate gaze distribution measures like gaze entropy [157], [181]. The same concept can be translated into VR eye tracking by analyzing transitions between OOIs. However, free head movement and limited precision and accuracy require careful data processing to obtain reliable OOI information. Depending on the size of the virtual objects, a participant can gaze at an object, but the measured gaze direction can miss the object. This makes a readjustment necessary during data processing, and some gaze targets must be estimated for further analysis.

To statistically analyze OOI information, the number of transitions and the OOI duration on specific objects can be used. However, these single values cannot reflect the complexity of gaze interactions. Networks are one way of representing the structure of gaze interactions with the OOIs. Networks provide several advantages that can be used to display and analyze gaze transitions between OOIs in virtual environments. Network analysis, grounded in mathematical graph theory [182], provides high scalability and can reflect network structures on

different levels of granularity [136]. It allows for the comparison of different networks (i.e., distribution measures), the investigation of connectivity within the network (i.e., interconnectedness measures), and the comparison of single nodes or groups of nodes (i.e., centrality measures) [183]–[185]. Modeling gaze data with networks is not only appealing because they provide good visualizations and interpretation, but research in cognitive science also suggests that networks can mimic the structure of the cognitive system and represent its dynamic processes [186], [187]. This makes the analysis of gaze networks especially interesting to investigate visual attention processes related to social and learning behavior. While there are some examples applying network analysis to gaze data [136], [145]–[148], [150], [188], applications and evaluations for the use of VR eye tracking data are sparse.

In summary, the research on eye tracking in virtual learning environments shows great potential to investigate information processing and learning. Well-established paradigms from eye-tracking research provide the basis for systematically investigating eye movements and visual attention in virtual environments. However, the specific characteristics of VR pose methodological and analytical challenges in modeling and interpreting VR eye-tracking data.

3. Research Objectives and Major Contributions

This dissertation aims to investigate how information processing and learning can be studied through eye-tracking analysis in virtual reality learning environments. Using VR as a reality simulator of a dynamic 3D environment, this thesis addresses the challenges of measuring information processing with eye tracking and the possibilities of modeling eye movements and visual attention. Given that VR can create a plausible and authentic experience of the learning situations while providing a standardized experimental setting, analyzing eye-tracking data can help to systematically address the following research questions:

[Q1] How do we encode and process information during the perception of 3-dimensional objects in VR? How can we reliably measure eye tracking during 3D scene perception?

[Q2] How do students distribute their visual attention in 3D virtual learning situations like a classroom? How do they focus on the lesson content and other learning-relevant events?

[Q3] How can eye-tracking information be modeled to analyze learning and social-related behavior toward virtual avatars in the virtual classroom?

Eye tracking, as a noninvasive measurement technique, is a promising assessment method since eye-tracking data is straightforward to obtain during VR experiments. Previous research in psychology, cognitive science, and human-computer interaction provides a sound body of literature on the analysis and interpretation. However, using eye tracking in VR also poses some challenges regarding integrating head movement, acquiring gaze target information, and interpreting eye movements in relation to information processing and learning.

This dissertation addresses some of the challenges of utilizing eye tracking in VR and proposes methodological and analytical solutions to these problems. This concerns obtaining reliable measures of pupil diameter, the use of gaze-ray casting to obtain the object of interest

information, network analysis, gaze entropy, and modeling eye movements using machine learning. Open-source tutorials and code are provided to allow other researchers and practitioners to utilize our solutions. The results of this thesis aim to expand the knowledge in the field of VR research in education science and explore the limits of analyzing and modeling eye tracking in VR. The results may not only provide insights for research on VR eye tracking and its connections to cognitive processing in virtual environments but could also help to strive towards effective virtual learning environments.

To further refine the three formulated research questions, the research articles presented in this thesis can be divided into three main contributions, representing the three sections in this chapter.

[C1] The first contributions in Section 3.1 present two standardized experimental testing environments in VR that focused on the processing and encoding of information with 3D objects and the reliable measurement of pupil diameter. The first testing environment asked participants to solve a mental rotation task, presenting pictorial 2D and visual 3D figures to compare performance and eye movements when solving the stimuli. The second testing environment was designed to collect pupil diameter values during a counting task to obtain reliable baseline measures.

[C2] Moving toward more authentic learning scenarios in Section 3.2, the classroom represents a rich learning environment that can be studied systematically to uncover learning-related processes. Different theory-driven design aspects of a virtual classroom were analyzed to understand how students distribute their visual attention in a virtual classroom. The results showed how manipulations in the classroom environment and different teaching events during the lesson affected students' visual attention.

[C3] In the last part in Section 3.3, gaze-based attention networks, obtained from gaze target information, were used to investigate students' visual attention patterns. The network approach using gaze transitions was evaluated on its utility for investigating the connection between visual attention distribution and social-related learning in the virtual classroom.

3.1. Information Encoding and Cognitive Load

3.1.1. The Impact of Presentation Modes on Mental Rotation Processing

This subsection is based on paper [1] from Chapter 1 *The Impact of Presentation Modes on Mental Rotation Processing: A Comparative Analysis of Eye Movements and Performance* published in *Scientific Reports* [1]. The full paper is presented in Appendix [1].

Motivation and Methodology

A fundamental aspect of information processing in virtual reality is the perception and encoding of three-dimensional objects. Since VR can create the perception of a 3D space, there are many ways to facilitate learning in cases where 2D representations cannot reflect all relevant information. For example, in situations where 2D representations can be ambiguous due to hidden or occluded parts. Further, recovering the structure of a 3D object from a 2D representation could cause additional effort during information encoding that might not be present for visual 3D objects. To create effective learning environments in VR, aspects of visual representations are relevant, and their effect on information processing and learning should be considered. On the one hand, there is great potential in creating visual 3D representations of learning material, whereas conventional materials only portray 2D representations of 3D objects. On the other hand, VR could cause additional processing or cognitive load due to its unique characteristics.

One way to investigate humans' perceptual encoding of 3D objects is to compare the performance and eye movements during a mental rotation task with pictorial 2D and visual 3D representations. Investigating mental rotation is especially suitable for various reasons. Mental rotation is a well-assessed construct from psychology that can be used to investigate spatial thinking and spatial reasoning. It has a strong and well-tested experimental design that can be used to assess participants' abilities. In addition, the eye movements of participants during solving mental rotation tasks have been analyzed in a number of previous studies. There is hardly any other task in psychology that has been studied more systematically with respect to individual processing steps and underlying cognitive processes. This gives researchers the opportunity to interpret the perception and information processing during this problem-solving task by analyzing eye movement data. Comparing participants' mental rotation performance and eye movements during mental rotation with pictorial 2D and visual 3D stimuli can be used to identify differences in visual information encoding and cognitive load.

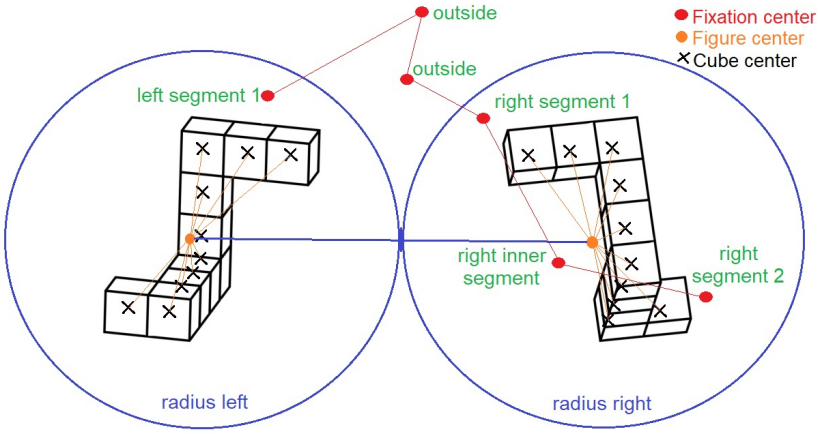


Figure 2.: An illustration of the procedure involved in obtaining the closest fixated segments for each fixation center.

For this reason, a mental rotation experiment was conducted in a virtual reality laboratory, where university students conducted two types of mental rotation tests each. One test presented pictorial 2D representations of 3D mental rotation stimuli displayed on a virtual screen in the VR laboratory. The other test presented visual 3D-rendered stimuli flowing above the experiment table in front of them. Participants' performance was tested in terms of the number of correctly solved stimuli and reaction time. The eye and head movements of participants were analyzed and compared to values of features previously identified as informative for stimulus processing and mental rotation strategies.

Since previous studies identified mental rotation strategies based on fixation patterns on specific parts of the figures, a primary goal was to obtain similar information. To identify fixations on specific parts of the figures and during head movements, first, a combination of a Velocity Identification Threshold (I-VT) and a Dispersion Identification Threshold (I-DT) algorithm was applied. The low accuracy and precision of the VR eye tracker only allowed for an estimation of the fixated figure segments. The fixated segments could be estimated by calculating the distance of the fixation center to the midpoints of the figure cubes and selecting the segments based on the shortest distance. This procedure allowed for an estimation of the fixated segment of the figures. With this information, different head, eye, and gaze features could be calculated. An illustration of this procedure is shown in Figure 2.

In addition to statistical tests, a Gradient Boosting Decision Tree (GBDT) classification algorithm was trained to identify the discriminative power of all head and eye-related features between the condition and to explore non-linear relationships in the data. Further, the

application of Shapley Additive Explanations (SHAP) could be used to obtain information on global and local feature importance.

Main Findings

In terms of mental rotation performance, participants solved significantly more stimuli correctly in the 3D condition with a significantly faster average reaction time. Notably, no sex difference was found for performance in either the 2D or 3D condition. This is consistent with previous findings that found no sex differences in mental rotation experiments when the experiment time was unlimited and more realistic representations were used. Regarding the different types of presented stimuli, participants made relatively more mistakes in the 3D condition with mirrored figures (unequal figures in which two segments of a figure are arranged mirrored to each other). They showed relatively longer reaction time for structural figures (unequal figures, in which one segment of one figure points in a different direction). A significant difference in eye and head movements between the conditions was found for features indicating differences in visual strategy, head movement, and cognitive load. With a total of 12 eye and head movement features, the GBDT algorithm was able to classify the two conditions with an average accuracy of 0.881 (SD = 0.011). This indicated that the feature contained relevant information for differentiating between the two conditions.

Based on the SHAP values and the results of the statistical significance tests, the following statements could be made about differences in the processing of pictorial 2D and visual 3D figures. According to the study of Xue et al., [189], specific eye movement patterns indicate different processing steps during the mental rotation process. In their study, the main differences were found between the first step of encoding and searching and the second step of transformation and comparison. Our experiment indicated that in the 2D condition, participants invested more time and effort in the first step of encoding and searching. This is consistent with the findings that processing visual 3D figures is easier than reconstructing a 3D representation from a 2D image. The results suggested that additional depth information in the 3D condition helped participants encode the visual figures faster and move to subsequent steps more quickly.

Further, in the 2D condition, participants had longer fixations on specific parts of the figures and lower saccade velocity, indicating a more focused exploration. Conversely, in the 3D condition, participants moved their heads closer to the figures, resulting in larger saccade amplitudes and higher saccade velocities. The increased pupil diameter in the 2D condition indicated greater perceived task difficulty for pictorial figures. Furthermore, the

presentation mode affected participants' strategies for solving the mental rotation tasks, with the 2D figures encouraging more piecemeal processing and the 3D figures more holistic processing.

However, faster encoding and more holistic processing in the 3D condition could have come with some drawbacks, indicated by more mistakes with mirrored stimuli and longer reaction times for structural figures. Specifically, the longer reaction time for structural figures suggested that participants took more time examining specific parts of the figure for this stimulus type in the 3D condition, potentially switching between holistic and piecemeal strategies.

Participants' better performance with visual 3D figures, faster encoding, and less cognitive effort indicated that 3D objects and 3D representations might also provide advantages in relation to learning materials and learning environments.

3.1.2. Pupil Diameter during Counting Tasks as Potential Baseline for Virtual Reality Experiments

While the obtained pupil diameter in the mental rotation experiment provided reliable measures due to the standardized experimental procedure, other VR experiments are characterized by a less structured environment in which participants are dropped directly into a lively situation. This causes problems in the processing of pupil information since, in order to measure the relative cognitive load of subjects in specific situations, the pupil values must be corrected to a baseline, which is more difficult to determine in these environments. In order to create the possibility of a reliable baseline measurement, the following experiment was carried out.

This subsection is based on paper [2] in Chapter 1 *Pupil Diameter during Counting Tasks as Potential Baseline for Virtual Reality Experiments* published in *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications* [2]. The full paper is presented in Appendix [2].

Motivation and Methodology

As introduced before, pupil diameter was found to be a reliable indicator of mental effort. An increase in pupil diameter was associated with an increase in mental effort [127]. This task-induced pupillary response has been observed in various cognitive tasks, including arithmetic [190], reading [191], and memory [128]. As such, it can indicate task difficulty and provide valuable insights into problem-solving and learning in VR, which are relevant

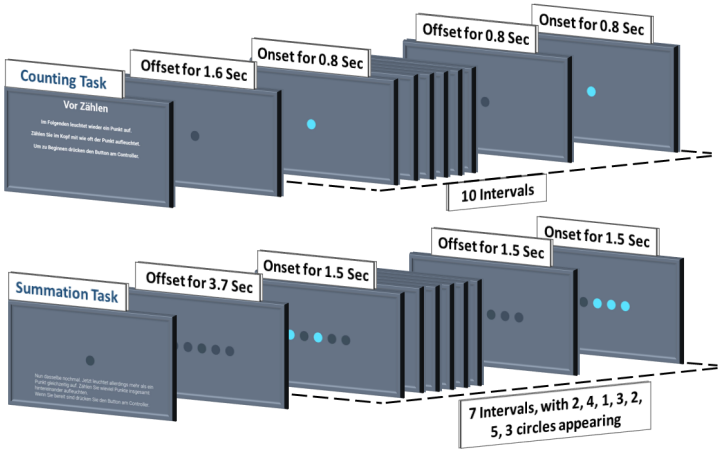


Figure 3.: Experiment design of the counting and summation task to measure pupil diameter baselines in VR.

aspects of education science. However, pupil diameter is idiosyncratic and must, therefore, be adjusted by a baseline to enable comparison across individuals [115]. In laboratory settings, a baseline can be established during a resting state [192] or a stimulus offset by staring at a black screen [115]. Although this procedure might be suitable for remote eye tracking, establishing an appropriate baseline for eye tracking in VR presents a more intricate challenge. When using an immersive VR with an HMD, establishing a proper baseline is challenging because exposing participants to a completely black screen may cause discomfort or fear [125], leading to confounded baseline measurements [122]. Additionally, variations in lighting levels [56] and participants' cognitive state, affected by emotional responses or mind wandering, may also influence baseline measurements [193]. To overcome these limitations, a VR environment with controlled visual conditions is needed to establish a reliable pupil diameter baseline measurement.

A short VR testing environment was created and evaluated to establish a pupil diameter baseline measurement for VR. Before and after the mental rotation task described in paper [1], the same participants conducted a counting and summation task in a separate VR environment without moving their heads. In the counting task, they were instructed to count the number of appearing dots in the middle of the screen, which was considered to be a low-demanding task. In the summation task, one to five dots appeared sequentially, and participants were instructed to sum up the total number of dots that appeared. The experiment design can be seen in Figure 3.

Main Findings

Pearson's correlation coefficients with $r > 0.855$ indicated significant consistency for pupil diameter in the counting task when correlating individual counting intervals (appearance of one dot). Similar correlations were found for the same experiment after the mental rotation task, but only acceptable retest reliability was found between both measurement times on the level of the individual intervals ($r < 0.7$). The pupil diameter significantly increased for the summation task, which showed that the additional task complexity in the summation task led to the expected increase in pupil diameter. Variations in pupil size during the tasks were mainly caused by the additional lightning induced by the dots appearing on the screen. However, when inducing the same lighting level during the counting task, the variation in pupil diameter values showed consistent patterns with similar high and low peak values for each counting interval.

These results highlight the potential of obtaining reliable pupil diameter measures in a separate testing environment before an experiment. However, the analysis of baseline-corrected pupil diameter values from VR eye trackers is only recommended as average values over longer time intervals. Due to the low temporal resolution of the VR eye tracker, analyzing more fine-grained pupillometry measurements (e.g., short amplitude changes) is not recommended [194]. Given the lower retest reliability, recalibrating the eye tracker for a longer VR experiment is suggested. This separates noise in the pupil diameter measure from other time effects, like fatigue or drowsiness [115].

Overall, one can conclude that this test procedure is time-saving, can be carried out quickly, and is only slightly influenced by factors such as head movements, different luminance levels, and mental states. By averaging during the counting task, a baseline can be calculated, and a subtractive or divisive baseline correction can be applied to control for idiosyncratic effects. However, this method cannot control for the effect of luminance on pupil size. This must be considered independently in addition to the idiosyncratic standardization proposed in this study.

3.2. Visual Attention in a Virtual Classroom

To investigate information processing and learning in a more realistic virtual environment, attention was drawn to virtual learning spaces, namely the virtual classroom. From the classroom experiment described in this section and in Section 3.3, three different evaluation studies could be conducted with the same data. The data was obtained from a virtual

(a) Cartoon-style virtual peer learners.

(b) Realistic-style virtual peer learners.

(c) Total classroom perspective from the back.

Figure 4.: Illustrations of the VR classroom with animated peer-learners and an animated teacher during a 15-minute lesson about computational thinking.

classroom experiment in which 381 sixth-grade secondary school students from Germany took part in this 15-minute VR lesson about computational thinking. A virtual teacher held the lesson and explained the content, referred to the learning content on the whiteboard (screen), and asked the students questions. In addition, 24 different students were individually animated in the classroom and participated in the lesson. Each participant in the experiment experienced the same lesson, but different aspects were manipulated. The participants were placed in different seating positions in the virtual classroom (second or last row), the presentation of the virtual characters was changed (cartoon-like or closer to real), and the participation and engagement of the virtual classmates were manipulated by varying their hand raising. In one of four scenarios, either 20%, 35%, 65%, or 80% of the virtual students raised their hand during teacher-student interactions. This resulted in a dataset with a $2 \times 2 \times 4 = 16$ between-subject design and different events or phases during the lesson (teacher explanation, teacher-centered discourse, questions, and answers), providing an even more fine-grained separation of the VR experiment data. Illustrations of the virtual classroom can be seen in Figure 4 (with hand-raising virtual students in cartoon style in Figure 4a, in realistic style in Figure 4b and a total perspective of the classroom from the back in Figure 4c).

3.2.1. Gaze-ray Casting

This subsection is based on the paper [5] *Gaze-based attention network analysis in a virtual reality classroom* published in *MethodsX* [5]. The full paper is presented in Appendix [5].

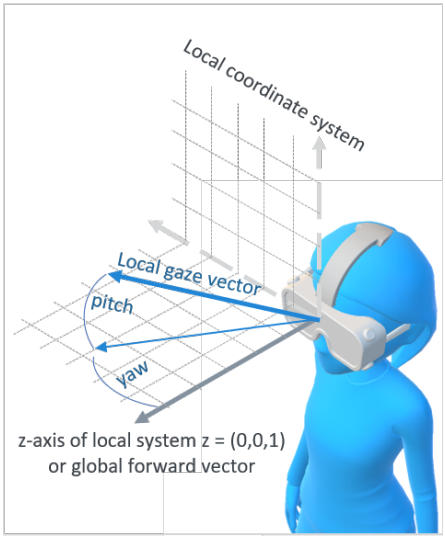


Figure 5.: Illustration of the calculation of the global gaze directions from local gaze directions using yaw and pitch rotation angles.

Motivation and Methodology

To obtain students' OOI information for the VR classroom experiment, which was modeled and analyzed in the two following articles, a software solution needed to be created for the Unreal Engine. This subsection describes the developed algorithm and its implementation in VR environments.

When analyzing the virtual classroom data, no software solution existed to obtain OOI information despite all head and eye tracking information being collected during the experiment. This meant that in an additional step, a data collection pipeline had to be developed that was able to perform gaze-ray casting in the Unreal Engine environment. The basic idea for gaze-ray casting already existed and was described in previous literature [179], [180], [195]. However, no software package was provided for the HTC Vive in combination with the Unreal Engine that automatically collected this information. To establish such a solution, an additional eye-tracking actor was created in the virtual environment that was aligned with the movement of the player and received the participants' local gaze direction of the eye tracker to transform it in real-time into the global gaze direction. This idea of creating one additional virtual actor had the advantage that one could easily implement the gaze-ray casting method into the already existing environment without interfering with the existing programming and game specifications.

The forward vector of the virtual actor, representing the participants' location in the environment, already pointed in the forward direction of the head, which meant that this vector only had to be adjusted according to the gaze direction to point exactly into the direction of the participants' eyes. The forward vector could be rotated by using the rotation angles yaw and pitch, which could be calculated from the local gaze directions. Since angle rotations are independent of the coordinate system, this procedure allowed for the calculation of the rotation angles in the local coordinate system of the eye tracker and could be used to rotate the global head direction vector.

To transform the local gaze direction into the global gaze direction, which could then be projected into the environment, basic Euclidean geometry was used. The calculation was based on the formula of calculating an angle α between two vectors A and B:

$$\alpha = cos^{-1}\left(\frac{A \times B}{|A| \cdot |B|}\right).$$

Yaw and pitch rotation angle were calculated by using the normalized local gaze vector $g = (x_g, y_g, z_g)$ and the coordinate transformed representation of the global forward vector $f = (0, 0, 1)$:

$$yaw = -cos^{-1}\left(\frac{z_g}{\sqrt{x_g^2 + z_g^2}}\right) \cdot \frac{180}{\pi \cdot sgn(x_g)}$$

and

$$pitch = cos^{-1}\left(\frac{x_g^2 + z_g^2}{\sqrt{x_g^2 + z_g^2}}\right) \cdot \frac{180}{\pi \cdot sgn(y_g)}.$$

The latter part of both formulas ensured the calculation of the angle in degree instead of radiant and adjusted for transformation from the Tobii eye tracker to the Unreal Engine coordinate system. An illustration of the vectors and angle in the local coordinate system is presented in Figure 5

The rotated forward vector, which then represented the global gaze direction, could be used as input for the ray-casting function, which was already implemented into the Unreal Engine function library. From the function output, the gaze target name, the gaze target location, and the distance to the gaze target could be obtained. A screenshot of the input and

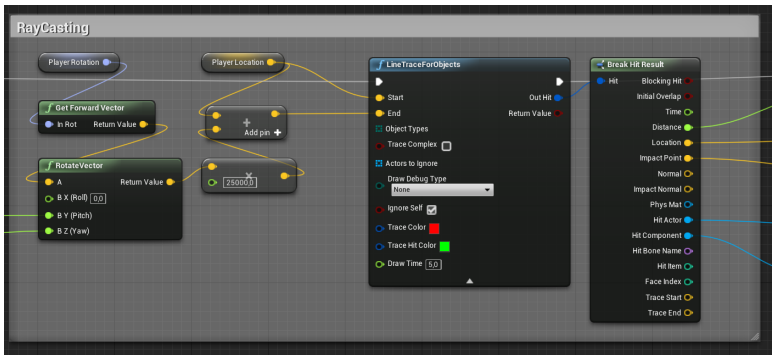


Figure 6.: Screenshot of the Unreal Engine blueprint displaying the ray-casting function with inputs and outputs.

output of the ray-casting function in the Unreal Engine blueprint can be found in Figure 6.

The code, as well as a tutorial of the implementation of gaze-ray casting with the Unreal Engine, was uploaded and made publically available on GitHub [1]. With this algorithm, the OOI information could be obtained, which could then be used in the following studies to analyze students' visual attention in the virtual classroom.

3.2.2. Students' Visual Attention in a Virtual Classroom

This subsection is based on paper [3] from Chapter 1 *Exploiting Object-of-Interest Information to Understand Attention in VR Classrooms* published in *2021 IEEE Virtual Reality and 3D User Interfaces* [3]. The full paper is presented in Appendix [3].

Motivation and Methodology

The analysis and modeling of visual attention using eye-tracking is task and environment-specific and, therefore, may not be transferable to other domains. Consequently, specific domain knowledge and configurations should be considered for the assessment of humans' visual attention in digital environments, especially in VR. To investigate how students distribute their visual attention in a virtual classroom, it is helpful to consider different configurations of the virtual environment.

Therefore, OOI information was obtained through gaze-ray casting with a focus on three main groups of objects. The aim was to analyze how long subjects ($N = 280$ from the full sample) gazed at the virtual classmates, the virtual teacher, and the screen and to understand

[1] `https://github.com/VRLabHIB/RayCasting_and_GazeBasedNetworkAnalysis`

the distribution of visual attention in this specific virtual classroom during the lesson. The OOIs were deliberately chosen as they are of particular interest in terms of attention to social dynamics and learning in a VR classroom. A full-factorial Analysis of Variance was conducted to investigate different gaze durations on all three OOIs for the different experiment conditions. This not only gave insights into the design of virtual classrooms (seating arrangement or visualization) in terms of the student's visual attention behavior, but it also simulated classroom behavior that might lead to implications for real-world learning scenarios (e.g., does the seating position affect visual attention).

Main Findings

Visual attention in terms of the total gaze duration on all three OOIs was investigated separately for the analysis. The total time spent on the virtual peer learners was larger when participants were seated in the back of the classroom, when they appeared in the cartoon style and in the 20% and 80% hand raising condition compared to the 65% condition. Reversed gaze duration was found for the total time spent on the virtual teacher and the screen. Participants spent more time on the virtual teacher when they were seated in the front and when the teacher and the virtual students were more realistic. Significantly more time was spent on the teacher in the 65% hand-raising condition compared to the 80% condition. While there was no difference in the time spent on the screen for the two visualization styles, participants in the front had significantly longer gaze durations on the screen. They also spend significantly more time on the screen in the 65% hand-raising condition in comparison to the 35% and 80% conditions.

This analysis showed that the students at the front of the classroom paid less attention to the peer learners and more attention to the learning-related content (teacher, screen). On the one hand, this is understandable, as the students at the front of the classroom had fewer peers in their FOV and had a clearer view of the teacher and screen. On the other hand, the differences in the hand-raising conditions, regardless of the seating position, showed that the students consciously focused their attention differently on the social cues from the peer learners. An interaction effect found between the hand-raising condition and sitting position showed that the influence of the sitting position was decisive for the distribution of attention to the virtual classmates. With regard to the design of effective virtual learning environments, one can conclude that if students should focus more on the content, then a computationally less complex, cartoon-like visualization of the virtual characters can be used. However, if a particular emphasis is placed on students' reactions to the social actors

(particularly in a collaborative learning environment), our results showed that the more realistic representations attracted students' attention.

3.2.3. Detect Classroom Discourse using Gaze Transition Entropy

This subsection is based on paper [4] from Chapter 1 *Using Gaze Transition Entropy to Detect Classroom Discourse in a Virtual Reality Classroom* published in *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications* [4]. The full paper is presented in Appendix [4].

Motivation and Methodology

The previous study investigated the VR classroom as a whole and focused on the overall effects of the different experimental manipulations for the full 15-minute lecture. When having a closer look into the VR classroom experience, the lesson was simulated to represent different phases of a lesson. To provide a learning experience that mimics traditional classroom teaching, the lesson included events of teacher explanation and events of student-teacher interaction. Certain events of classroom discourse [40], [41] in a teacher-centered lesson have been simulated, such as teacher questions toward the class, hand raising by the virtual peer learners, and answers to the questions by singular virtual students. The active participation of the virtual students and the interaction of the teacher with the students provided some elements of classroom discourse. The aim of this study was to find out if participants' visual behavior could indicate different events during the lesson [196]. While during teacher explanations, the main focus should be drawn to the front, during discursive events, participants should switch their attention towards all social actors in the classroom and actively explore the student-teacher interactions [15], [57].

For this, gaze measures could be utilized to indicate the distribution of visual attention and the extent of visual exploration. Stationary and gaze transition entropy [157], [181], [197] was analyzed to investigate the difference in participants' visual behavior in the two main events during the lesson (teacher explanation vs. elements of classroom discourse). The two entropy measures were calculated for 30-second intervals with a sliding window of 10 seconds to reflect dynamic changes in gaze entropy over time. The transition matrices consisted of all peer learners, the teacher, and the screen as separated OOIs. Given the predefined animations in the VR classroom, the events could be distinguished according to a timetable. Further investigation indicated that gaze entropy measures could also distinguish the extent of visual exploration for the different hand-raising conditions. The explanatory

value of both entropy measures was studied by two multilevel linear regression models with the events and the hand-raising conditions as independent dummy variables. To further test the predictive power of both entropy measures for discerning classroom events, a logistic regression model was trained to predict the event using only the two entropy measures as independent variables.

Main Findings

Transition entropy was significantly higher during events of classroom discourse compared to events of teacher explanations. Moreover, transition entropy was higher in the 20% and 80% hand-raising condition, compared to the category of average hand-raising (the combination of the 35% and 65% condition). Stationary entropy was also significantly higher for classroom discourse events and significantly higher in the 80% condition compared to the average category. This indicated that participants showed more visual exploration during the events of teacher-student interactions. Visual exploration was particularly strong for participants in the 80% and 20% hand-raising conditions. Notably, one can assume that these conditions provide relevant information for social comparison and learning (everyone or no one is participating or knows the answer). The 20% hand-raising condition showed a significant effect for transition entropy but not for stationary entropy. This suggested that although participants engaged in more visual exploration, they only spent time on a few hand-raising students. Interestingly, there was no interaction effect between the events and the hand-raising conditions. This indicated that the entropy values were not only larger during the discursive events, but participants showed generally higher values during the full lesson. It is also possible that there were variations over time that averaged out the interaction effect, e.g., a different level of visual exploration at the beginning than at the end of the scene.

The clear distinction between the events in terms of gaze entropy indicates that the students' showed the intended attention to their virtual classmates in situations where social-related learning information could be obtained from the lesson. Since visual exploratory behavior varied even between conditions, it indicated that even simple hand-raising signals mimicking engagement and participation are perceived in the virtual classroom. On the one hand, these results highlight the impact of social characters in the virtual learning environment. On the other hand, the successful application of gaze transition entropy to differentiate visual attention behavior highlights the value of analyzing gaze transitions to detect social-related behavior.

3.3. Gaze-based Networks and Learning with Simulated Classmates

One article in this section investigated the same virtual classroom experiment. In contrast to the previous studies, it focused on the relationship between visual attention and learning-related outcomes like interest, self-concept, and performance. Gaze-based attention networks were exploited to analyze the students' visual attention in the classroom, and structural network measures could be associated with the learning outcomes. Alongside this article, a corresponding methods article was published to explain the eye tracking analysis using the network approach. This methodological article is presented first.

3.3.1. Gaze-based Attention Network Analysis

This subsection is based on the second part of the paper [5] *Gaze-based attention network analysis in a virtual reality classroom* published in *MethodsX* [5]. The full paper is presented in Appendix [5].

Motivation and Methodology

By using gaze-ray casting in VR, one can directly obtain the OOI information and calculate the transitions between the different OOIs. This procedure offers an alternative approach to calculating fixations and saccades, which is more difficult to obtain in 3D virtual environments. Assuming that a gaze transition to a certain object also means the visual processing of this object, the frequency and sequence of gaze transitions can provide indications regarding information processing. As already shown in Section 3.2.3, distribution measures built from transition information indicate visual exploration toward social actors. Instead of using single measures, scan paths are alternative representations of gaze transitions [198], [199]. One way to represent scan path information is to count the number of transitions between all selected OOIs over a period of time and build a transition matrix where one matrix entry represents the number of transitions from one specific OOI to another. While this transition matrix is used to calculate entropy measures [157] or to train machine learning classifiers [114], it also represents a directed graph [182], [200], [201].

The gaze transition matrix can be seen as an interaction-based network in which the OOIs are the nodes, and the number of transitions are the weighted edges of the network. All gaze transitions of one participant during the VR experiment can be collected in one network. This network reveals a participant's visual attention distribution in the virtual

scene. It opens up the possibility of analyzing the network structure of one participant by calculating different structural variables and using them to compare the structure between participants. Structural variables are measures that indicate the centrality of specific nodes in the network, the distribution of transitions within the network, and their connectedness. The evaluation of the gaze-based attention network analysis with regard to social-related behavior and learning is described in Section 3.3.2. The technical and methodological aspects of gaze-based attention networks and how to calculate structural variables are described in this section.

In the case of the classroom experiment, the transition matrix was obtained for each participant for the full 15-minute virtual lesson. An illustration of an example network can be seen in Figure 7. Specifically, for the purpose of investigating learning with simulated classmates as described in paper [6] [6], structural variables were calculated that could extract network information about the visual attention towards the virtual peer learners and the learning material. The runtime performance of the pipeline was also evaluated, and recommendations regarding data storage could be formulated.

Main Findings

Several structural variables of the networks turned out to be feasible for analyzing participants' visual attention in the virtual classroom. In general, three different types of variables could be calculated to describe the structure of the gaze-based attention networks.

Centrality measures state the importance or prominence of a node or a set of nodes. It describes the weighted number of connections a node holds to all other nodes. Degree centrality, in particular, sums the weights of all incoming and outgoing edges, which translates to the number of gaze transitions towards and away from one OOI. The degree centrality of a set of nodes is calculated by only counting the incoming and outgoing edges from outside the set. This means that degree centrality portrays the importance of an OOI in terms of visual attention.

Distribution measures describe the distribution of gaze transitions among specific nodes or across the whole network. Participants' gaze transitions could either occur between a small set of OOIs or between multiple OOIs, which would be indicated by different distribution values. Weighted degree centrality was the first distribution measure [183] that was used to describe the gaze distribution for single nodes. Contrary to what the name suggests, this is a measure for calculating the distribution of all outgoing edges of one node. A simplified calculation and an adaptation of this method for unsorted networks were presented. As a

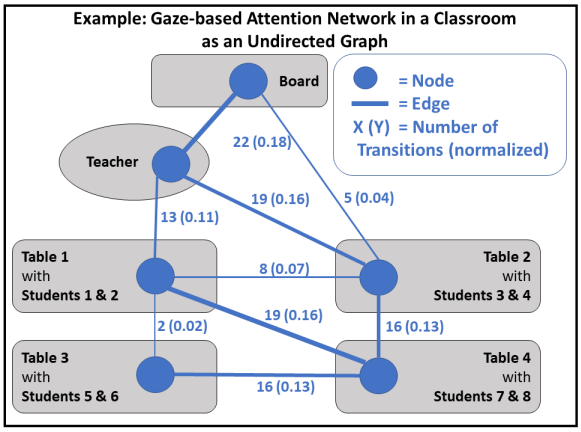


Figure 7.: Example for a gaze-based attention network of the virtual classroom with fewer nodes and edges.

second measure of distribution, the uniformity of the edge weights was calculated using chi-square statistics. The uniformity measure was applied to the whole network structure, indicating equal or unequal distributions of the gaze transitions between the OOIs.

The last type of structural variables represented the interconnectedness within the network. These measures were particularly useful in separating visual attention towards the teacher and screen and towards the virtual students in the VR classroom. Cut size was used as a measure to count the number of weights between two subgraphs representing the teacher and screen or virtual students. Further, the concept of cliques in networks was used to investigate the frequency of gaze transitions among virtual students. A clique in a network represents a subset of nodes that are all connected with one another. Therefore, the size and number of cliques represent well-connected substructures in the network. If participants transitioned their gaze fully between a subset of virtual students, this group was considered a clique and represented a more intense visual exploration of this subgroup of virtual students.

Performance analysis of the data pipeline showed that the small number of variables required to calculate the networks and the sequential execution of the individual calculation steps greatly reduced the size of the data files and the computational runtime. Calculating and storing networks instead of data frames further reduced the size of the data. With specific data formats from network packages, large networks can be stored in a more space-efficient manner. The possibility of calculating the structural variables for each participant individually avoids more computationally intensive methods for comparing networks.

3.3.2. Learning with Simulated Virtual Classmates

This subsection is based on paper [6] from Chapter 1 *Learning with simulated virtual classmates: Effects of social-related configurations on students' visual attention and learning experiences in an immersive virtual reality classroom* published in *Computers in Human Behavior* [6]. The full paper is presented in Appendix [6].

Motivation and Methodology

The classroom can be understood as a central learning environment. A regular classroom environment is characterized by a few structural elements. There is a whiteboard (screen) with the learning material, a teacher, and a peer learner who participates in the lesson. The social interactions and social relationships between all classroom actors create a dynamic of mutual learning, which contributes to students' emotional, cognitive, and academic development [90]. Although the social interactions of the virtual avatars are predefined, they can still reflect social behavior in real classroom situations to a certain degree [91]. VR, in particular, is capable of simulating the complex and dynamic learning processes in a classroom and incorporating various contextual and peer-related factors. Learning in a social context is of importance for educational science and influences the (perceived) learning achievement and motivation of students [202], [203].

For this reason, this study focused primarily on students' visual attention toward virtual peer learners and its relation to specific learning outcomes like interest, self-concept, and achievement collected in a pencil-paper post-test. To model students' visual attention, gaze-based attention networks were computed for each participant, and specific structural variables were calculated (as described in Section 3.3.1) and statistically compared with the learning outcomes. Further, the structural variables were also investigated with regard to the experimental conditions of the experiment (sitting position, visualization style, and hand-raising).

The structural variables that were calculated focused specifically on differentiating between the teacher, the screen, and the peer learners and, more precisely, distinguishing the visual distribution of attention among the virtual peer learners. Degree centrality was calculated for the teacher, the screen, and the subset of nodes containing all peer learners. Furthermore, the number of cliques containing only peer learners and their average size was calculated. To investigate further if participants showed different behavior regarding specific virtual peer learners, the proportion of girls to boys in the cliques of peer learners was counted. To analyze the distribution of visual attention in the virtual classroom, the weighted degree centrality of

the screen, the cut size between teacher/screen and peer learners, and the uniformity for all gaze transitions were calculated.

Main Findings

Comparable with the previous study [3], the structural variables showed a clear tendency of the students to focus more on the teacher and screen when seated in the front. Further, students in the back showed higher numbers in all clique and distribution-related variables. The visualization condition showed less consistent results. However, the majority of peer learner-related variables were higher for the cartoonish avatars. Participants showed a higher degree centrality toward the teacher and screen and a lower degree centrality toward peer learners in the medium 65% hand-raising condition, compared to the 20% and 80% conditions. These results showed that not only did the gaze duration on the OOIs differ for the experimental condition, but also that participants observed the virtual peer learned with different frequencies and intensities.

Partial correlation between the structural variables and the learning-related outcomes revealed that participants' self-reported interest was particularly associated with differences in visual attention behavior. The degree of centrality of peer learners, the number and size of the cliques, and the proportion of boys in cliques were negatively associated with participants' interest in the lesson. These results revealed that students who were more interested in the lesson content also focused less on the virtual peer learners and more on the screen, as indicated by the positive association between interest and degree centrality on the screen. Students' average situational self-concept (which consisted of four items like: "I could solve the robot tasks faster than the others") was negatively associated with the proportion of boys in the cliques. The post-test score of students' achievement showed a positive correlation with the degree centrality on the screen, indicating that students who focused more on the learning content also showed better learning.

Overall, the study's findings revealed significant associations between students' visual attention distribution in the virtual classroom and learning-related outcomes. They also showed that gaze-based attention networks analyzed through structural variables can reveal information about learning with social avatars in virtual learning environments.

4. Discussion

This thesis contributes to the investigation and understanding of information processing and learning in virtual reality learning environments. The exploitation of VR eye-tracking data showed great potential in analyzing learning-related aspects of information encoding, visual attention in learning environments, and social aspects of learning. The results presented in this thesis could be a first step towards understanding the mechanisms of human (visual) behavior in VR in order to strive for the effective use of VR as a learning environment.

Both methodological and theoretical aspects were addressed. With regard to the analysis of eye movement data in VR, several methodological contributions could be achieved. More specifically, a standardized test environment was used to show how a reliable measurement of pupil diameter could be achieved in order to obtain physiological parameters relevant to learning, such as the change in the pupil for estimating cognitive effort. In addition, a reliable procedure was established to obtain gaze target information as a combination of head and eye movements in virtual environments of the Unreal Engine using the gaze-ray casting method. Providing an implementation tutorial with open-source access to the code allows other scientists to replicate this method and can, therefore, be used for further scientific research. Furthermore, a method was developed and evaluated to enable the assessment of different aspects of visual perception when applying network analysis to VR eye-tracking data. From the successful application of the gaze-based network analysis, two conclusions could be drawn. Network structures can model the complexity of visual attention to some degree, especially when the connections between specific OOIs are meaningful. Further, structural variables cannot only determine network structures in the sense of meaningful representations, but they are also easy to interpret and allow for statistical comparison. While some methodological contributions could be presented in this thesis, no systematic methodological evaluations were carried out, as the contributions in this thesis focused more on their application.

Theoretical implications for the effectiveness of VR for learning could be drawn from the investigation of information encoding with 3D objects, the investigation of visual attention

in a virtual classroom, and the reaction of students toward social cues and social-related learning behavior during the virtual lesson. It could be shown that in a standardized mental rotation experiment, subjects showed a better mental rotation performance with visual 3D objects. Eye movements indicated easier and more holistic processing and suggested that subjects preferred more egocentric processing of 3D figures by changing perspective.

Depending on the sitting position and the visualization style of the virtual avatars, students showed different visual attention behaviors. They focused more on the teacher and screen when sitting in the front and on the teacher when a more realistic visualization style was used. This further demonstrated that design aspects clearly influenced students' visual attention in the classroom and provided further evidence that the virtual classroom situation introduced behavior also expected in a real classroom. Further, when taking a closer look at the structure of the lesson, the lesson design evoked the expected behavior of the participants. They focused more on the teacher and learning content during events of teacher explanation while shifting their attention towards the virtual avatars during elements of classroom discourse. Further, the different levels of student engagement manipulated by the hand-raising conditions, even more, showed the sensitivity of participants toward learning-related social cues in the virtual classroom.

Further, this thesis highlighted the potential of utilizing different representations of visual attention in the form of networks or gaze entropy to investigate the social-related behavior of participants. Children's strong reaction toward social avatars was not unexpected. Already, early experiments in psychology showed a high anthropomorphic tendency of humans to anticipate human emotions and personality traits when watching animated movies of triangles and circles [204]. Our tendency to personify our environment and perceive objects in the context of social behavior and norms opens up a great potential to create virtual social situations in VR. This also reduces the costs of conducting research on social behavior in VR since rendering realistic social avatars is more difficult, and these animations often suffer from an uncanny valley effect [205]. It seemed not to be necessary for participants that the virtual avatars were particularly real to show tendencies of real social interactions.

4.1. Limitations

Despite the contributions of this thesis, there were also various limitations when analyzing eye movements and visual attention information using VR eye-tracking. For example, in the mental rotation experiment [1] (paper [1]), the spatial resolution of the VR eye tracker

did not allow a fine-grained analysis of the gaze targets. Even though the stimulus material was placed right in front of the participants, most of the obtained gaze target points missed the figures. A detailed scan path analysis of the fixations on single cubes of the figures and a precise tracing of gaze patterns was not possible. The low precision and accuracy in the eye tracking data could have been compensated by additional information. Eye tracking is only one source of information that explains information processing and learning in humans. While this thesis contributes to the explanation of human visual behavior and its relation to cognitive processing, other physiological measures or language information were not considered in the analysis.

Furthermore, all the VR environments studied in this thesis had in common the fact that the locomotion of the participants was limited, and the virtual scene showed only a few dynamic movements. Although this might limit the generalization of the eye-tracking results to other VR experiments, it is also an accurate representation of most current real-life learning environments. Because one intention of education science is to uncover the learning process in real-life situations, the presented VR environments must also represent these situations. From the perspective of eye tracking, however, the generalization of these results must be tested in scenes with more dynamic moving entities. As stated before, the analysis of visual attention through eye-tracking is very scene and knowledge-dependent and can, therefore, not be generalized across different learning scenarios. Specifically, obtaining reliable gaze target information might be even more challenging when objects move quickly in the environment. However, a positive side effect of these environments with fewer movement possibilities was that participants did not show motion sickness. Because motion sickness is usually a result of the disparity between visual and vestibular stimuli [16], [206], fewer head and body movements lower its occurrence.

4.2. Virtual Reality in Education Science

From the perspective of empirical psychological research, VR seems an appealing research tool because it allows research that escapes the traditional experiment situation while still allowing for standardized experimental testing.

However, VR research might face the same situation as multimedia research in general. Randomized control studies are, from a research design perspective, the most valid experimental setup to investigate differences systematically. However, their feasibility can be questioned when it comes to the multitude of possible combinations for designing effective

VR environments. In contrast to treating the VR environment as a black box and interpreting only differences between the treatment and control groups, the investigation of participants' behavior during the experiment can be one step towards establishing explanations and predictions of human behavior in VR environments that can lay the foundation of a theory of human behavior in VR. This might even become more important when the possibilities to interact and manipulate the virtual scene increase.

For example, the results from the mental rotation experiment [1] indicated that participants switched from an object-based transformation to an egocentric transformation for visual 3D figures. Especially for the processing of spatial relations, the change of perspective can play an important role and correspond better or worse with the participants' usual mental manipulation of 3D objects [207]. While the visual 3D figures made this perspective change possible, which was not possible with pictorial stimuli, other aspects were neglected. For example, an additional reduction of cognitive load could also be achieved by the physical rotation of the objects. The interplay of motor and cognitive skills, in particular, which could be provided in interactive virtual environments, could be the next step to understanding further aspects of the learning process in VR [208], [209].

In order to create virtual learning environments that are not only effective from a design point of view but also provide insights for learning research outside of VR, it is necessary to clarify how learning in VR differs from learning in non-virtual environments.

Generally speaking, most of this thesis's results stem from the interdisciplinary exchange between computer science, education science, and cognitive psychology. One way to approach the pending question of the effectiveness of VR for learning could be to integrate different perspectives from different fields. These interdisciplinary endeavors could help establish a broader theory of effective learning in VR. However, many aspects remain to be uncovered on the path toward a theory of human behavior in VR.

4.3. Virtual Reality in Education Practise

From the experiences made during data collection and backed up by further evidence from other literature [65], design aspects, clear instructions, and intuitive usability are key elements to make VR easily accessible for practitioners. However, while there are already many VR environments from different software companies, the pedagogical and learning-oriented aspects are often overlooked. There are plenty of studies on the effectiveness of VR for learning, but evidence from research on education showed that VR is rarely used in practice

[210].

While most of this thesis's findings concentrated on using VR in education science, some implications can be formulated for its use in practice. Displaying 3D objects should result in a better understanding of their form and shape and, therefore, allow for faster encoding, better understanding, and reduce students' cognitive load. This means that we should create virtual learning situations and materials, especially where real-life representations are hard to construct. This can be either when there are limitations in scale (e.g., exploring the stars in the universe), limitations in visualizations (e.g., multivariate Gaussian Distributions), or when learning opportunities are sparse (e.g., for medical operations).

4.4. Diversity of Learning Situations

Conventional education environments, such as a classroom or traditional learning materials, like a school book, create a specific learning situation. In relation to Foucault's idea of an apparatus dispositif [211], these elements shape and determine the learning situation and affect the learner's behavior. This also implies that specific students could benefit from a given learning situation or from the mode of presentation of the learning material, while others could show difficulties. This principle is not limited to just a few individual differences, yet the results of the studies in this thesis have provided particular insights into sex differences.

In traditional mental rotation tests with 2D representations, large sex differences were found in test performance. This led to the formulation of various implications on the spatial ability between sexes in psychological literature [212]–[215]. Interestingly, the VR experiment was not able to reproduce these sex differences in mental rotation performance in our experiment, which is in line with other experiments presenting mental rotation in VR [216], [217]. While the result of a non-significant sex difference was to be expected for the visual 3D figures, there was no difference for the pictorial 2D figures as well. When comparing the experimental design with a traditional mental rotation test, it stood out that an overall time constraint to solve the mental rotations task was missing. This also emphasizes the importance of the testing situation and how it can affect individuals differently. Further, the study results showed the importance of considering multiple representations to help students who have trouble encoding 2D representations from learning materials. In this case, VR showed great potential in providing the additional dimension necessary. This especially highlights that the findings from the experiments, which were meant to assess specific abilities, were also shaped and determined by their experimental design. Although

this insight is by no means new, it is still a valuable lesson to keep in mind. This should encourage researchers to think about the different interacting factors when creating learning and testing environments and to consider diversity in learning scenarios.

Furthermore, the findings regarding students' situational self-concept in the classroom study [6] indicated that the sex-specific composition of the virtual class and the distribution of attention to virtual avatars with differently-read sex characteristics were important for students' perceived competence. Especially in subject-specific learning situations (like the presented lesson on computational thinking) with certain gender-stereotypical beliefs, the diversity aspect with regard to virtual social avatars is even more important.

Given the ongoing discussion about the effectiveness of VR for learning, the more simple conclusion might be that VR can at least create more diverse learning situations and materials. While some situations and materials might impose more difficulties for some students, they might facilitate learning for others. So, while there might not be a VR environment that increases learning for everyone, it still creates new possibilities to approach learning from a different perspective, which may be more suitable for the specific needs of a particular learner. Therefore, VR has great potential to diversify the landscape of learning and empower individuals with different abilities.

4.5. Towards Effective Virtual Reality Learning Environments

As an outlook on this thesis, we can return to the statement that an effective learning environment comprises many different elements. To make VR an effective tool for learning, the design of the VR environment, the cognitive mechanisms during learning in VR, and the monitoring of the learning process must be interlinked.

Although learning environments such as the virtual classroom are particularly relevant for research, e.g., for assessing learning processes, this severely limits VR's possibilities. Rigorous experimental designs and the investigation of quasi-realistic learning scenarios ignore the value of exploration, curiosity, and self-efficacy that VR environments can offer. Future research could aim to bridge this gap and increase the learner's possibilities in VR through better measurement of learning processes. For example, gaze target information could be used in real-time in the VR environment to influence the behavior of virtual avatars. Students who are being looked at could behave differently, or the teacher could intervene if participants are not paying attention. This would enable research into adaptive behavior in the virtual

classroom.

This thesis presented different models and methods for analyzing eye-tracking data. Nevertheless, further research is necessary to explain the underlying mechanisms of information processing and learning. Analyzing eye movement and gaze-related features can certainly provide some insights into the learning mechanism. However, this does not account for the complex interactions between different cognitive processes and probably misses important patterns between eye movements and behaviors. While gaze-based attention networks can model gaze transition information, future research should expand on the capabilities networks and incorporate more node and edge information from different sources. The incorporation of multiple sensory information and the use of graph neural networks might be one approach to building computational models that predict participants' behavior. Multimodal data assessment, in combination with methods from computational modeling, might provide a promising opportunity to bridge the gap between the complex nature of human behavior and the limited information researchers are able to obtain. Future research could specifically concentrate on the development of models that are dynamically able to predict participants' behavior in virtual environments based on the sensory information collected.

Despite the great potential of analyzing human behavior via eye tracking, there are also ethical considerations to be taken into account. When having access to rich behavioral information, protecting the privacy of persons is inevitably important. Eye tracking data can, for example, be used to obtain personal information like age, gender, or health [45], [218], and therefore, privacy issues have to be handled appropriately [219]. Specifically, VR eye-tracking information collected from students and children could not only be used to improve learning but also for commercial reasons. With future research heading towards more elaborate models of visual attention and a more detailed analysis of the learning process [220], different approaches of privacy-preserving methods must be considered [59], [218], [219], [221], [222].

With regard to design aspects in VR, a number of developments can be expected as a result of the increased use of generative AI. This technology will make it increasingly easier to create your own VR environments and generate more realistic representations [223]. There is also great potential for education science in terms of the possibility of interaction with objects and social avatars in the environment. The flexible creation of and interaction with virtual learning environments can mean a further leap forward for the use of VR in educational practice. Especially in the context of adaptive learning, generative models can be a promising way of dynamically interacting with a VR environment.

These further aspects provide us with an idea of the potential of VR to investigate information processing and learning in future research and the versatile opportunities for developing effective learning environments. While there is a great interest in VR from the research perspective, there are fewer applications of this technology in actual practice. A great deal of effort is still needed to translate the empirical findings, including those in this paper, into practical applications.

A. Information Encoding and Cognitive Load

The following publications are enclosed in this chapter:

[1] **P. Stark**, E. Bozkir, W. Sójka, M. Huff, E. Kasneci, and R. Göllner, "The impact of presentation modes on mental rotation processing: A comparative analysis of eye movements and performance", *Scientific Reports*, 2024. DOI: 10.1038/s41598-024-60370-6

[2] **P. Stark**, A. Tobias, O. Milo, and K. Enkelejda, "Pupil diameter during counting tasks as potential baseline for virtual reality experiments", in *2023 Symposium on Eye Tracking Research and Applications (ETRA '23)*, Germany: ACM, Jun. 30, 2023, p. 7. DOI: 10.1145/3588015.3588414

A.1. The impact of presentation modes on mental rotation processing: A comparative analysis of eye movements and performance

A.1.1. Abstract

Mental rotation is the ability to rotate mental representations of objects in space. Shepard and Metzler's shape-matching tasks, frequently used to test mental rotation, involve presenting pictorial representations of 3D objects. This stimulus material has raised questions regarding the ecological validity of the test for mental rotation with actual visual 3D objects. To systematically investigate differences in mental rotation with pictorial and visual stimuli, we compared data of $N = 54$ university students from a virtual reality experiment. Comparing both conditions within subjects, we found higher accuracy and faster reaction times for 3D visual figures. We expected eye tracking to reveal differences in participants' stimulus processing and mental rotation strategies induced by the visual differences. We statistically compared fixations (locations), saccades (directions), pupil changes, and head movements. Supplementary Shapley values of a Gradient Boosting Decision Tree algorithm were analyzed, which correctly classified the two conditions using eye and head movements. The results indicated that with visual 3D figures, the encoding of spatial information was less demanding, and participants may have used egocentric transformations and perspective changes. Moreover, participants showed eye movements associated with more holistic processing for visual 3D figures and more piecemeal processing for pictorial 2D figures.

A.1.2. Introduction

Mental rotation, the ability to rotate mental representations of objects in space, is a core ability for spatial thinking and spatial reasoning [224], [225]. Mental rotation is required for everyday skills, like map reading or navigating, and is an important prerequisite for individuals' learning [226]. Higher mental rotation performance is associated with higher fluid intelligence and better mathematical thinking [227]. It has been found to be beneficial for students' learning in mathematics domains such as geometry and algebra [228]. Thus, mental rotation ability acts as a gatekeeper for entering STEM-related fields in higher education [229].

A standardized test by Shepard and Metzler [230] for measuring humans' mental rotation performance displays two-dimensional (2D) images of two unfamiliar three-dimensional

(3D) figures. For these pictorial stimuli, participants are instructed to determine whether the two figures are identical. For this, the two figures are depicted from different perspectives by independently rotating one of them along its axis [230], [231]. Individuals' performance in mental rotation is reflected by the number of correct answers and task-solving speed (reaction time) [232], [233]. Since its initial development, this experiment has been replicated many times [212], [233]–[235]. The test by Shepard and Metzler is one of the most frequently used tests to examine mental rotation. It laid the foundation for understanding spatial cognition [132], [236]–[238] and continues to be referenced in contemporary research [233], [239], [240]. Replicating this classic experiment allows researchers to build on a well-established foundation and examine enduring principles of mental rotation.

However, its ecological validity to assess real-life mental rotation has been questioned [241], [242]. Developments in the field of virtual simulations enable experiments to be conducted with increased ecological validity yet still under controlled and standardized conditions [16]. In particular, virtual realities (VR) have become powerful tools in psychological research [6], [85]. VR allows for the creation of environments with 3D spatial relations that can be explored and manipulated by users and are experienced in an immersive way [10]. This allows for the presentation of visual 3D figures, rendered as 3D objects in the environment, and introduces visual and perceptual differences to pictorial (2D) stimuli.

The pictorial stimuli of the conventional mental rotation test are orthographic, parallel representations of 3D figures on a planar surface (as images). This pictorial representation lacks two sources of depth information present in visual (3D) figures when placed in a VR environment with realistic spatial relations [47]. The first source of depth information is provided by stereoscopic vision due to binocular disparity. The binocular disparity stems from the slight offset between the two displays projected onto the two eyes in the head-mounted display (HMD), enabling stereopsis and depth perception [243]. This depth cue is particularly relevant for 3D vision, where it contributes to participants' ability to perceive depth and spatial relationships between objects. The second source of depth information is introduced by motion parallax [47], [244]. Motion parallax, also known as structure-from-motion, emerges as a consequence of real-time head tracking and rendering based on the observer's position within the virtual space. This dynamic depth cue allows users to perceive the 3D structure of objects by moving their heads. As they move relative to the 3D object, the representation of the object is updated and provides different views to identify the object. Furthermore, shadows provide additional depth information. They occur when physical objects interact with light sources in a VR environment. Shadows contribute to the

perception of object volume and spatial relationships in visual figures. Presenting mental rotation stimuli in VR provides the most comprehensive visual information. In contrast, rear-projection systems offer solely pictorial information [216], and stereoscopic glasses introduce binocular disparity [245], leaving motion parallax as the final piece of the puzzle added by VR [217].

This additional visual information is expected to affect participants' stimulus processing and mental rotation strategy when solving items with visual stimuli in comparison to pictorial representations. A series of processing steps when solving mental rotation tasks have been identified [131], [189]: (1) encoding and searching, which combines the perceptual encoding of the stimulus and the identification of the stimulus and its orientation; (2) transformation and comparison, which includes the actual process of mentally rotating objects; (3) judgment and response, which combines the confirmation of a match or mismatch between the stimuli and the response behavior.

One would expect the visual modes of presentation to introduce differences in the processing steps. During encoding and searching with pictorial figures, a model of the 3D object structure must be recovered from a planar 2D representation [246]. This reconstruction process has been found to be a demanding task [247] and should not be necessary with visual figures. One would also expect the identification of the stimulus and its orientation to be more demanding with pictorial figures. A displayed image remains static regardless of the observer's location; therefore, participants have to make assumptions about occluded or ambiguous parts of the figure. For pictorial figures, the additional head movement might even produce perceptual distortions described by the differential rotation effect [248], in which the size and shape of images are perceived inappropriately when the observer is not in the center of the projection [249]. In contrast, binocular disparity and motion parallax would constantly update the visual 3D figures based on the participants' relative location to the object. Test takers can explore the visual figures and gather additional information from different perspectives, which should help them to identify the figures and their orientation more easily.

In the second step of transformation and comparison, mental rotation involves manipulating and rotating mental representations of geometric figures in the mind. Exploiting motion parallax with visual 3D figures could reduce the need for extensive mental transformations. For example, participants could reduce the rotation angle between the figures through lateral head movement. The rotation angle is the degree to which the figures are rotated against each other. This may make the comparison process more intuitive and less cognitively demanding.

Motion parallax due to head movement could also lead to a shift from the object-based transformation of the stimuli to an egocentric transformation [207]. In object-based transformations, the observer's position remains fixed while the object is mentally rotated. An egocentric transformation involves a change of perspective, rotating one's body to change the viewpoint and orientation. It has been found that egocentric transformations, as a form of self-motion, are more intuitive and result in faster and more accurate mental rotation [208].

Similar reaction times for mental and manual rotation suggest that participants mentally align the figures to each other for comparison [250]. Two prominent alignment strategies have been described for mental rotation: piecemeal and holistic. The piecemeal strategy involves breaking down the object into segments and mentally rotating the pieces in congruence with the comparison object to assess their match. A holistic approach entails mentally rotating the entire object and encoding comprehensive spatial information about it [251], [252]. In their original study, Shepard and Metzler viewed the linear relationship between rotation angle and reaction time as evidence against conceptual or propositional processing of visual information [230], [253]. Later research, which investigated the process of rotation itself, revealed that both a holistic and a piecemeal approach were used to align the figures [132], [252], [254]. When processing visual figures, motion parallax allows for lateral head movements, which could be used to decrease the rotation angle between the figures by changing perspectives. The additional depth information due to binocular disparity could facilitate the comparison of spatial relationships between object features. These aspects might enable a more holistic processing of the figures.

Regarding judgment and response, participants are expected to perform better with visual 3D figures than with pictorial 2D figures. Lower cognitive demands during encoding might result in faster stimulus processing. The potential to apply an egocentric transformation and more holistic processing can be expected to lead to more efficient and more accurate responses with visual 3D figures.

The process of mental rotation is reflected in eye movements, which capture the visual encoding of spatial information [189], [235]. Eye movement metrics can provide comprehensive information on stimulus processing and mental rotation strategies [77], [235], [251], [252], [255], [256]. Basic experiments have shown that eye movements are controlled by cognitive processes, and consequently, it is possible to distinguish task-specific processes [119]. For example, different mental rotation strategies were identified and discriminated based on fixation patterns derived from eye-tracking data [132]. Fixation measures that incorporate spatial information are expected to reveal relevant information about stimulus processing.

Different fixations on different segments of the figures have been associated with the first or second processing steps [189]. During the step of encoding and searching, the majority of fixations targeted one segment of one figure, whereas, in the second step of transformation and comparison, fixations targeted all segments of both figures equally. This should lead to a higher fixation duration on singular segments in the first step and an equal fixation duration on all parts of the figure in the second step.

Saccadic movements between fixations, measured by saccade rate or saccade velocity, have also been utilized to investigate mental rotation with pictorial figures [77], [189], [257]. Directional saccadic movements containing spatial information can reveal temporal dependencies in stimulus processing [189]. For example, a backward saccade that guides the eye toward a previous location is called a regressive saccade [258]. We would expect that the regression towards a previous location could either be a need for information retrieval of figure information or a back-and-forth between congruent figure segments during the comparison step.

Regarding mental rotation strategies, information about the number of transitions between figures compared to the number of fixations within the figures has been applied to quantify the use of holistic vs. piecemeal strategies [77], [252]. The ratio of the number of within-object fixations divided by the number of between-objects fixations has been shown to indicate holistic processing (ratio ≤ 1) or piecemeal processing (ratio > 1) [252], [259].

The pupil diameter provides information about the size of the pupil in both eyes and can be used to detect changes due to contraction and dilation. An increase in pupil diameter has been associated with higher cognitive load [127], [128], [260], as the Locus Coeruleus (LC) controls pupil dilation and is engaged in memory retrieval [261], [262]. Moreover, two different measures of pupil diameter behavior have been attributed to the phasic and tonic modes of LC activity [261]. Tonic mode activity is indicated by a larger overall pupil diameter and is associated with lower task utility and higher task difficulty. Phasic mode activity is indicated by larger pupil size variation during the task and is associated with task engagement and task exploitation [235], [263]. While solving mental rotation tasks, a larger average pupil diameter over individual trials could indicate tonic activity, whereas a larger peak pupil diameter as a task-evoked pupillary response could indicate phasic activity [235], [262].

Recently available devices for analyzing eye movements in VR experiments include eye-tracking apparatuses. These devices record sensory data frame by frame to track visual and sensorimotor information in a standardized way during experiments [264]. The VR's HMD additionally allows for tracking head movement. Changes in head movement serve

as a valuable indicator of whether participants make use of motion parallax. A recently published study by Tang et al. [77] analyzed eye movements during a mental rotation task in VR, but solely for visual 3D figures. The results of their VR experiment showed that the mental rotation test with visual 3D figures replicates the linear relationship between rotation angle and reaction time. Lochhead et al. [217], on the other hand, investigated performance differences between pictorial and visual 3D figures presented in VR. Their results indicated that participants exhibited higher performance in the 3D condition compared to the 2D condition. However, they did not use eye tracking to capture participants' visual processing of the stimuli to potentially explain presentation mode effects on performance.

Our study used a VR laboratory (see Figure A.2) to examine individuals' mental rotation performance for pictorial 2D figures and visual 3D figures with the Shepard and Metzler test. We examined eye and head movements from N = 54 university student participants to determine differences in stimulus processing and mental rotation strategies when solving mental rotations with pictorial and visual stimuli. In both conditions, 28 stimuli pairs were shown, modeled after the original figures by Shepard and Metzler [230]. In the 3D condition, stimuli were rendered on a virtual table in front of the participants, allowing them to view the figures from different perspectives by moving their heads. In the 2D condition, the stimuli appeared on a virtual screen placed on the table at the same distance from the participants as in the 3D conditions. A series of 3D and 2D figures were presented, with the two conditions randomized block-wise within each student. For each task, participants' performance in terms of the number of correct answers and reaction time as well as eye-movement features were recorded. The following hypotheses were formulated:

First, we expected participants' performance in solving mental rotation tasks to be better with visual 3D figures than with pictorial 2D figures. Second, we expected the visual differences to evoke differences in stimulus processing and mental rotation strategies, which may indicate differences in performance between the two modes of presentation. To investigate this hypothesis, we analyzed how eye and head movements differed during task-solving in both conditions. To ensure that we could compare all stimulus pairs between the two conditions, no overall time limit was set for the experiment.

In addition to utilizing statistical analysis, we implemented a Gradient Boosting Decision Tree (GBDT) [265] classification algorithm to identify the experimental condition based on eye and head movements. This machine learning approach surpassed traditional linear statistical methods, which are often limited to linear relationships between features and the target variable. Successfully predicting the experiment condition based on eye and head movement

features would demonstrate the importance of these features for the distinguishing task.

Behavioral data, such as eye and head movements, are characterized by temporal dependencies and determined by biological mechanisms (e.g., a fixation is followed by a saccade and vice versa), which often results in high collinearity between the features [266]. From the class of machine learning models, we selected GBDT rather than other models like Support Vector Machines or Random Forest because of its ensemble approach. Ensemble methods can handle some degree of collinearity by partitioning the feature space into separate regions [176]. Previous research has demonstrated the suitability of GBDT models for spatial reasoning tasks involving geometrical objects, which are comparable to the task utilized in this study [177].

Provided that the GBDT model classifies the conditions correctly, a Shapley Additive Explanations (SHAP) explainability approach can be applied [178]. The SHAP approach provides information on both global and local feature importance. Global feature importance ranks input features by their significance for accurate model predictions, identifying the most relevant features for differentiating between the experimental conditions. Local feature importance supplements this by providing additional information on the relationship between feature variables and target variables. It reveals which feature values were attributed to each condition and how effectively those values distinguish between conditions. These aspects complement statistical analyses and offer valuable insights into the relationship between eye movements and mental rotation processing.

A.1.3. Results

Mental rotation performance differences

All participants completed both experimental conditions (2D and 3D) in a block-wise randomized condition order. The mean values and standard deviations of all variables in each condition are depicted in Table A.1. Further information about the distributions is presented in Supplementary Table S1. We used a non-parametric, paired Wilcoxon signed-rank test since some variables were not normally distributed. We report the Z statistics from two-tailed, paired tests with p-values. Additionally, we applied a two-tailed, paired t-test and compared the results for skewed distributions (Supplementary Table S2).

On average, participants spent 11.91 minutes in VR (SD=3.65 minutes) without any breaks in between. In the 2D condition, participants solved 83.2% of the stimuli correctly on average ($M = 0.832$, $SD = 0.105$), while in the 3D condition, they solved 88.2% correctly ($M = 0.882$,

Feature	**2D** ($M \pm SD$)	**3D** ($M \pm SD$)
Percentage solved correctly	0.832 ± 0.105	0.882 ± 0.101
Reaction time (s)	6.861 ± 3.583	6.076 ± 3.214
Mean fixation duration (s)	0.218 ± 0.025	0.216 ± 0.028
Mean fixation rate (n/s)	2.239 ± 0.266	2.301 ± 0.32
Mean regressive fixation duration (s)	0.142 ± 0.051	0.177 ± 0.042
Equal fixation duration between figure (ratio)	0.695 ± 0.086	0.721 ± 0.079
Equal fixation duration within figures (ratio)	0.187 ± 0.065	0.449 ± 0.084
Strategy ratio ($\lessgtr 1$)	1.488 ± 0.948	0.77 ± 0.292
Mean saccade velocity ($°/s$)	239.186 ± 20.838	250.476 ± 22.439
Mean saccades rate (n/s)	2.016 ± 0.466	2.151 ± 0.451
Mean pupil diameter (mm)	0.039 ± 0.095	-0.096 ± 0.123
Peak pupil diameter (mm)	0.314 ± 0.101	0.416 ± 0.104
Mean distance to figure (cm)	88.599 ± 8.584	86.567 ± 10.21
Mean head movement to the sides (cm)	4.942 ± 3.595	5.713 ± 3.438

Table A.1.: Mean values and standard deviations were aggregated on the participant level separately for each dimension ($n = 54$). Units are either seconds (s), number per second (n/s), a ratio between 0 and 1, or greater and smaller than 1 ($\lessgtr 1$), angle in degrees per second ($°/s$), millimeters (mm), centimeters (cm), or centimeters per second (cm/s).

$SD = 0.101$). Participants achieved a significantly higher percentage of correct answers in the 3D condition ($Z = 243$, $p = .001$) when comparing the 2D with the 3D condition in a two-tailed test. Participants exhibited a longer reaction time (in seconds, $M = 6.861$, $SD = 3.583$) in the 2D condition than in the 3D condition ($M = 6.076$, $SD = 3.214$). Based on a two-tailed test, reaction time differed significantly between the conditions ($Z = 1168$, $p < 0.001$). Details of the statistical analysis are shown in Table A.2.

To ensure that the differences in performance could not be attributed to sex differences, we performed additional statistical analyses to verify this. No sex differences were found in our study. This is consistent with previous research, which reported no sex differences in experiments conducted without time constraints [234], [239] or using less abstract stimulus materials [235], [267]. Detailed statistics can be found in Supplementary Table S3.

We verified that the performance differences between 2D and 3D are not attributed to order effects. The average reaction time was always found to be higher in the 2D condition, regardless of the order. However, the differences were larger when the 2D condition was presented first. Similar results were observed for the percentage of correctly solved stimuli,

Feature	Z	P	M diff	95% CI	Effect size
Percentage solved correctly	243	0.001	-0.071 ± 0.014	$[-0.089, -0.036]$	-0.550
Reaction time (s)	1168	< 0.001	0.644 ± 0.2	$[0.317, 1.02]$	0.573
Mean fixation duration (s)	873	> 0.999	0.003 ± 0.003	$[-0.002, 0.008]$	0.176
Mean fixation rate (n/s)	504	0.48	-0.064 ± 0.029	$[-0.125, -0.002]$	-0.321
Mean regressive fixation duration (s)	113	< 0.001	-0.034 ± 0.005	$[-0.045, -0.024]$	-0.848
Equal fixation duration between figures (ratio)	477	0.276	-0.023 ± 0.01	$[-0.043, -0.003]$	-0.358
Equal fixation duration within figures (ratio)	1	< 0.001	-0.265 ± 0.011	$[-0.286, -0.244]$	-0.999
Strategy ratio ($\lessgtr 1$)	1384	< 0.001	0.642 ± 0.106	$[0.446, 0.868]$	0.864
Mean saccade velocity ($°/s$)	160	< 0.001	-11.568 ± 1.692	$[-15.208, -7.671]$	-0.785
Mean saccade rate (n/s)	339	0.012	-0.148 ± 0.035	$[-0.215, -0.07]$	-0.543
Mean pupil diameter (mm)	1438	< 0.001	0.134 ± 0.014	$[0.104, 0.164]$	0.937
Peak pupil diameter (mm)	38	< 0.001	-0.099 ± 0.011	$[-0.121, -0.079]$	-0.949
Mean distance to figure (cm)	1253	< 0.001	1.313 ± 0.681	$[0.682, 2.114]$	0.688
Mean head movement to the sides (cm)	230	< 0.001	-0.618 ± 0.215	$[-0.911, -0.368]$	-0.69

Table A.2.: Wilcoxon signed-rank tests comparing the 2D and 3D conditions ($n = 54$). P-values of all eye and head features were Bonferroni-corrected to account for multiple comparisons. A positive median difference value indicates a higher median value in the 2D condition ($\pm$ standard error). The 95% confidence interval for the median difference and rank biserial correlation effect size is reported. Units are either seconds (s), number per second (n/s), a ratio between 0 and 1, or greater and smaller than 1 ($\lessgtr 1$), angle in degrees per second ($°/s$), millimeters (mm), centimeters (cm), or centimeters per second (cm/s).

for which the main differences were only present if the 2D condition was presented first. We also ensured that the sexes were equally distributed in both groups. The respective descriptive statistics can be found in Supplementary Table S4. In order to ensure that mental rotation in VR replicates expected differences, we provide additional descriptive statistics regarding reaction time and rotation angle for each condition separately in Supplementary Table S5.

To test for potential interaction effects between the experimental condition and the stimulus type (equal, mirrored, and structural), we conducted a multi-level regression analysis for each performance, eye, and head feature as the independent variable with condition and stimulus type as categorical independent variables. All analysis results and a model description can be found in Supplementary Table S10. Compared to equal figures, mirrored figures revealed a significantly lower percentage of correctly solved trials for the 3D condition. Structural figures, compared to equal figures, showed a significantly longer reaction time in the 3D condition.

Statistical differences in eye and head movements

We tested for differences in all eye and head movement features between the two conditions using two-tailed, paired Wilcoxon signed-rank tests with aggregated values on the participant level. To consider multiple comparisons, all reported p-values were Bonferroni-corrected before.

Regarding fixation-related features, we found no significant difference in the mean fixation duration ($Z = 873$, $p > 0.999$) and the mean fixation rate ($Z = 504$, $p = 0.48$). However, the mean fixation duration following a regressive saccade differed significantly between the conditions ($Z = 113$, $p < 0.001$), with a higher duration in the 3D condition than in the 2D condition. The feature equal fixation duration between the figures showed no significant difference ($Z = 477$, $p = 0.276$) after correcting for multiple comparisons. The feature equal fixation duration within the figures showed a significant difference, with an equal distribution in the 3D condition ($Z = 1$, $p < 0.001$). The strategy ratio comparing the number of fixations within and between the figures showed a higher mean value for the 2D condition ($Z = 1384$, $p < 0.001$).

Regarding saccade-related features, there was a significant difference in mean saccade velocity ($Z = 160$, $p < 0.001$), with a higher mean value in the 3D condition. A higher mean saccade rate was found for the 3D condition ($Z = 339$, $p = 0.012$). Mean pupil diameter showed significantly higher values in the 2D condition ($Z = 1438$, $p < 0.001$), while peak pupil diameter was significantly lower in the 2D condition ($Z = 38$, $p < 0.001$). The mean distance to the figure and mean head movement to the sides differed significantly with closer distances to the figure in the 3D condition ($Z = 1253$, $p < 0.001$) and larger head movement to the sides in the 3D condition ($Z = 230$, $p < 0.001$).

Regarding the interaction between the experimental condition and the stimulus type, three features showed significant interaction effects. When correcting for multiple comparisons, equal fixation duration within the figure showed lower values in mirrored figures (compared to equal ones) in the 3D condition. For structural figures (in comparison to equal ones), participants showed a higher mean saccade velocity and a lower mean saccade rate in the 3D condition (see Supplementary Table S9 and S10).

GBDT model capabilities

We trained a GBDT model to predict the experimental condition at the level of individual trials based only on eye and head movement features. 80% of the data was used for training, with a random train-test split. In 100 iterations, predictions for the test set exhibited an

average accuracy of 0.881 (with $SD = 0.011$). The best-performing model had an accuracy of 0.918. False classifications were balanced between the two target conditions, with 27 trials misclassified as the 2D condition and 22 misclassified as the 3D condition. A confusion matrix for the best-performing model predictions is given in Table A.3.

	2D labeled	**3D labeled**
2D predicted	267	22
3D predicted	27	280

Table A.3.: Confusion matrix for 596 predicted trials (classified as either 2D or 3D) in the test set. Predictions of the best-performing GBDT model out of 100 iterations with a random 80 : 20 train-test split.

Explainability results

We applied the SHAP Tree Explainer [178] to the best-performing model. Equal fixation duration within the figure was rated the most important feature for the GBDT model, with smaller values leading to predicting the 2D condition and larger values the 3D condition. The second most important feature was mean pupil diameter, with a higher mean pupil diameter leading to predicting the 2D condition. The third most important feature was the strategy ratio, with higher values leading to predicting the 2D condition and low values the 3D condition. Peak pupil diameter was identified as the fourth most important feature, with the opposite tendency as mean pupil diameter. A higher peak pupil diameter led to predicting the 3D condition. Mean distance two the figure (5th) showed a tendency to predict the 2D condition for higher values. However, there is higher variability in feature values in both conditions. For the following three features, mean regressive fixation duration (6th), mean saccade rate (7th), and mean head movement to the sides (8th), the model showed a tendency to associate higher values with the 3D condition. The remaining features exhibited little importance for model prediction or no clear tendency towards one condition or the other. The results are visualized in Figure A.1. Based on the additional analysis for multi-collinearity (see Supplementary Table S6), we found no high correlations between the individual features. A larger negative correlation was found between mean saccade rate and mean fixation duration ($r = -0.39$) and between mean saccade rate and strategy ratio (-0.31).

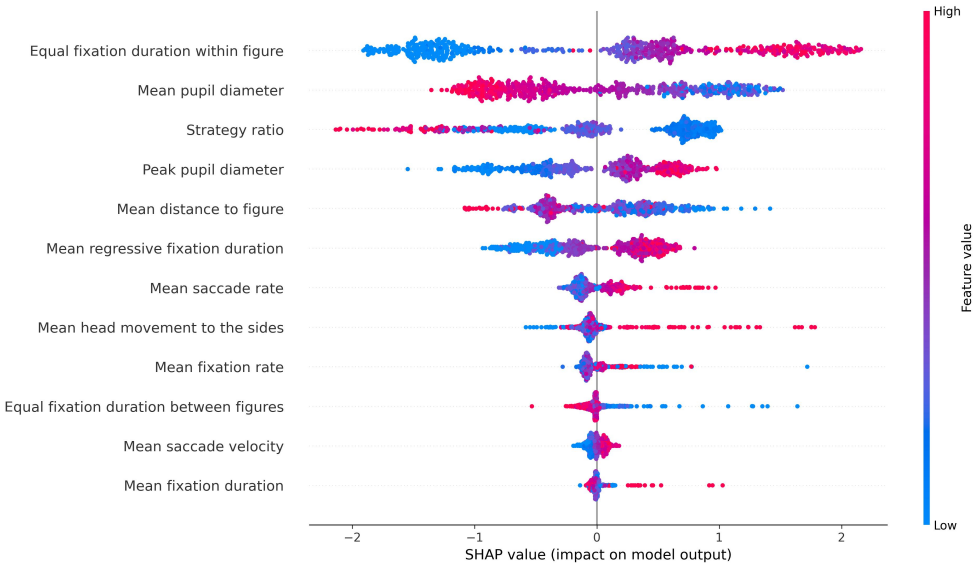


Figure A.1.: Summary plot of SHAP values for the GBDT model with the best performance out of 100 iterations (accuracy 0.918). Features are ordered according to their importance for the model's predictions. The x-axis describes the model's prediction certainty towards 2D (left side) and 3D (right side). Data points are predicted trials. The red color indicates that the data point has a high value for the feature, and the blue color indicates that the data point has a low value for that feature

A.1.4. Discussion

This study used a VR laboratory to test mental rotation, presenting Shepard and Metzler [230] stimuli in a controlled yet ecologically valid environment. Specifically, our study investigated whether the mode of presentation (i.e., pictorial 2D or visual 3D figures) evoked differences in visual processing during task solving and affected participants' performance. Participants' mental rotation test performance differed significantly between the two presented conditions, with higher accuracy and shorter reaction time in the 3D than in the 2D condition. These findings are in line with previous research reporting better performance for 3D figures [213], [217]. We argued that the direct encoding of visual figures would allow for faster and easier processing in the 3D condition, leading to a decrease in response time. In addition, we argued that access to depth information via binocular disparity and motion parallax would enhance stimulus perception and facilitate the transformation and comparison of visual figures. These

factors could have led to improved performance on mental rotation tasks in the 3D condition. In addition, motion parallax in the 3D condition provided the opportunity to use head movements to change perspective (e.g., egocentric perspective taking). In combination with easier perception of the geometric structure of the figures, this could have led to a more holistic processing of the stimuli.

We analyzed eye and head movement information to substantiate these assumptions. We argued that the changes introduced by the mode of presentation and their effect on stimulus processing and mental rotation strategies can be investigated by analyzing participants' visual behavior. The successful training of the GBDT model indicated that the eye and head movement features provided valuable information to distinguish between the two conditions. Statistical analysis, as well as SHAP values, discriminated different eye and head movement patterns in both conditions.

Overall, our results indicate that the additional information provided by motion parallax led to more pronounced head movement to the sides and a closer inspection of the visual 3D figures. In turn, directly inspecting hidden parts of the depicted figures by changing perspective could have resulted in a less ambiguous perception of the figure [268].

At a more detailed level, our findings suggest that fixation patterns in the 2D condition related more strongly to the first processing step of encoding and searching, while patterns in the 3D condition were related to the step of transformation and comparison. Xue et al. [189] found that the first step was associated with more fixations on particular segments of the figures. In contrast, the second step showed a more equal distribution of fixations across all segments of the figures. The SHAP value analysis indicated that the two conditions mostly differed in fixation duration within the figures. A less equal distribution within the figures, which implies longer fixations on particular segments, was found in the 2D condition. This supports the claim that the availability of depth information through motion parallax and binocular disparity accelerated the initial encoding of the visual figures and allowed participants to move more quickly to subsequent steps. In the same vein, a lower saccade velocity was found in the 2D condition, indicating more saccades within particular segments of the figures. However, in the 3D condition, participants moved their heads, on average, closer to the figures. This increases the saccade amplitude since the distances between and within figures become larger, which in turn increases saccade velocity [269]. The inverse correlation of -0.24 between saccade velocity and distance to the figure indicates that, at least to some degree, saccade velocity is affected by participants' head movements (see Supplementary Table S6).

Furthermore, the mean pupil diameter was larger in the 2D than in the 3D condition, while the peak pupil diameter was smaller in the 2D condition than in the 3D condition. The larger mean pupil diameter as an indicator of tonic activity could imply higher task difficulty and lower task utility in the 2D condition. This can be further supported by the lower saccade rate in the 2D condition. A decreasing saccade rate was previously associated with an increase in task difficulty [270]. In contrast, the smaller peak pupil diameter as an indicator of phasic activity could imply lower engagement and less task-relevant exploitation of the 2D task. These results provide further evidence that the first step of encoding might be more demanding for the pictorial 2D figures, and additional information due to head movement might have facilitated task-relevant exploitation. Moreover, a shorter average fixation duration after a regressive saccade in the 2D condition could indicate a need for more information retrieval when trying to maintain a 3D mental model of the figures in mind.

At the same time, our study findings indicate that presentation mode might confound previous research on individuals' strategies for solving mental rotation tasks. The presentation of 2D figures was more strongly related to features indicating a piecemeal strategy than the presentation of 3D figures. This was implied by differences in the strategy ratio used to distinguish between holistic and piecemeal strategies [247], [252]. Our results showed that participants in the 2D condition moved their gaze more frequently within a figure and switched fewer times between figures than in the 3D condition. Consequently, one might assume that the 2D presentation mode could evoke piecemeal processing. In this case, however, the strategy ratio not only reflected the way in which the figures were compared but could also be affected by differences in the first step of encoding the figures. Our results clearly speak to the relevance of different processing steps, which need to be considered more carefully in future research. For instance, the reason why mental rotation seems to be easier with more natural stimuli [267] could be that encoding figure information is less demanding.

Results of the interaction analysis indicated that a faster encoding of the figure and more holistic processing in 3D were associated with some costs. Participants made relatively more mistakes with mirrored stimuli in the 3D condition, and took a relatively longer time for structural figures compared to equal figures. In addition, eye movement features showed that participants took more time investigating specific parts of the figure for structural stimuli compared to equal stimuli in the 3D condition. When searching for the misaligned segment in structurally different stimuli, participants potentially switched from a holistic strategy to a piecemeal strategy, which in turn resulted in longer reaction time with this stimulus type.

In sum, our study showed how eye and head movements could be used to investigate systematic differences in stimulus processing and mental rotation strategies across different modes of presentation. However, we are also aware of the potential limitations of the present study. Although we were able to show that the mode of presentation causes a difference in processing, we cannot determine, for example, in which of the steps individuals with high and low abilities differ. Furthermore, our results suggest that the strategies used are related to the mode of presentation. Although we identified strategies using a common indicator [247], [252], future studies should expand on this using more elaborate methods, such as ones allowing for time-dependent analyses. Moreover, the accuracy of the VR eye tracker was a technical limitation of our study. Previous studies using the same eye-tracking device have reported lower gaze accuracy in the outer field of view [169]. By using the VIVE Sense Eye and Facial Tracking SDK (Software Development Kit) to capture eye-tracking data in the Unreal engine, the frame rate of the eye tracker was adjusted to the lower refresh rate of the game engine. Therefore, our eye tracking in VR did not provide the same spatial and temporal resolution as remote eye trackers. There was also a limitation regarding the usability of head-mounted displays (HMD). Although we used the latest VR devices in our experiment, the participants had the added weight of the HMD on their heads, and we had to connect the HMD device to the computer with a cable. This limited the participants' freedom of movement to some degree and may have affected the extent of their head movement and natural exploration. Another limitation concerns a possible confounding effect between head movement and fixations due to the vestibular eye reflex. This reflex stabilizes vision when fixating during head movement and could, therefore, compromise fixation-related features due to the influence of automated adjustments [271], [272]. The bivariate correlations between -0.07 and 0.11 revealed only small relationships between both head movement and all fixation-related features for both the 2D and 3D conditions on the level of individual trials (see Supplementary Table S7 and S8). While one cannot rule out the effect of vestibular eye reflex on fixation-related features, the study findings indicated a similarly small influence of the vestibular eye reflex on fixations in both conditions.

Despite these limitations, VR proved to be a useful tool to test mental rotation ability in an ecologically valid but controlled virtual environment. We made use of integrated eye tracking to learn more about the impact of presentation modes on stimulus processing and mental rotation strategies when solving Shepard and Metzler stimuli. Our results indicated that mental rotation places different demands on different processing steps when processing pictorial or visual figures. The demands that pictorial 2D figures place on participants, from

encoding to rotating the figures, seem to be ameliorated by the provision of additional visual information. More importantly, our results suggest that 2D figures evoke piecemeal analytic strategies in mental rotation tasks. This, in turn, leads to the question of whether piecemeal processing tells us more about the ability to create and maintain 3D representations of 2D images than it does about the ability to rotate one 3D figure into another.

A.1.5. Methods

Participants and procedure

During data collection, 66 university students participated in the experiment. Due to missing eye-tracking data, we had to exclude 12 participants. Data from 54 participants remained for the analysis. In the remaining sample, 33 participants stated their sex as female and 21 as male. Participants' average age was 24.02 (SD = 7.24), and 35 of them needed no vision correction, while 19 wore glasses or contact lenses.

The experiment took place in an experimental lab at a university building. After providing written informed consent to participate, participants completed a pre-questionnaire. The pre-questionnaire asked for socio-demographic and personal background information. Before using the VR, participants were informed about the functionality of the device and a five-point calibration was performed with the integrated eye tracker. After that, participants conducted the mental rotation test in VR. In the test, participants had to go through 60 stimuli one after another. Each stimulus displayed two Shepard and Metzler figures, for which participants had to respond whether they were equal or unequal using the handheld controllers [230]. 30 of the stimuli were presented on a virtual screen, replicating a classical computerized Shepard and Metzler test (2D condition). The other 30 stimuli were displayed as 3D-rendered objects floating above a table (3D condition). Participants were randomly assigned to first see all 2D or all 3D stimuli. Randomization was used to balance out any kind of sequence effect. Out of the 54 participants, 31 saw the 2D experimental condition first, and 23 saw the 3D experimental condition first. No time limit was set for completing the tasks. After completing the experiment, participants received compensation of 10€. The total experiment did not exceed one hour, and the VR session did not exceed 30 minutes. To complete both VR conditions, participants spent, on average, 11.91 minutes in VR (SD=3.65 minutes) without any breaks in between. The study was approved by the ethics committee of the Leibniz-Institut für Wissensmedien in Tübingen in accordance with the Declaration of Helsinki.

Experiment Design

VR environment

The VR environment was designed and implemented in the game engine Unreal Engine 2.23.1 [273]. Participants sat on a real chair in the experiment room and entered a realistically designed virtual experiment room, where they also sat on a virtual chair in front of a desk (see Figure A.2). Before the start of the mental rotation task, instructions were shown in the 3D condition on a virtual blackboard located behind the experimental table in the participants' direct line of sight, whereas for the 2D condition, the instructions were presented on the virtual screen display. Participants were instructed to solve the tasks correctly and as quickly as possible. Additionally, participants completed one equal and one unequal

Figure A.2.: Images taken from our VR environment show the virtual experiment room as well as example stimuli from the 2D and 3D conditions embedded in the environment.

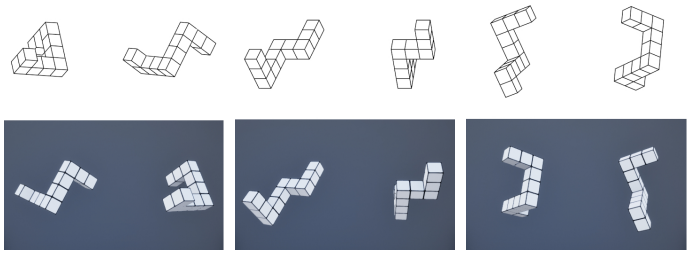
example stimulus pair, after which they received feedback on whether the examples were solved correctly or incorrectly. After they responded with the controllers, a text was displayed on the blackboard or the screen. The stimuli appeared at a distance of 85 cm from the participants. For the 2D condition, the stimulus material appeared on a virtual computer screen placed on the desk. During the 2D condition, the screen was visible at all times; only in the center of the screen did the figures appear and disappear. In the 3D condition, the stimulus material appeared floating above the table. The 3D figures were rendered as 3D objects in the environment, which allows the figures to be viewed from all perspectives. The distance to the center of the 3D figures was the same as the distance to the screen in the 2D condition. The figures were also placed at the same height in both conditions. Before a stimulus appeared, a visual 3-second countdown marked the start of the trial. Participants then decided whether figures were equal or unequal and indicated their response by clicking the right or left controller in their hands (left = unequal, right = equal). Instructions on using the controllers were displayed on the table in front of them.

Stimulus Material

Our mental rotation stimuli were replications of the original test material by Shepard and Metzler [230]. The 2D mental rotation test was designed as a computerized version and presented on the VR virtual screen. For the immersive 3D condition, the original test material was rendered as 3D objects in VR. In both conditions, each stimulus consisted of two geometrical figures presented next to each other.

One figure was always a true-to-perspective replication of the Shepard and Metzler material used in previous experiments [213], [274]. These figures and their form of presentation have been used in various studies and provide a reliable and valid basis for our experimental material [225], [235], [275], [276]. These stimuli were created by rotating and combining ten base figures [277]. Each base figure was a 3D geometrical object composed of 10 equally sized cubes appended to each other. The cubes formed four segments pointing in different orthogonal directions. This resulted in three possible combinations for the figure pairs: Either they were the same (equal pairs) or not the same (unequal). If unequal figure pairs had the same number of cubes per segment, but one figure was a mirrored reflection of the other, we called it an unequal mirrored pair. If the unequal figure pairs were similar, except one segment pointed in a different direction, we called it an unequal structural pair. Examples for all three stimulus types are depicted in Figure A.3. Variation in task difficulty was induced by rotating one figure along its vertical axis by either 40, 80, 120, or 160 degrees while keeping the other figure in place. Ergo, each stimulus showed one of the four rotation angles. Due to incorrect visual displays, two stimuli had to be removed from the experiment since different figures were presented in the two conditions. This resulted in 28 stimuli used for data analysis. For all 28 stimuli, we ensured a relatively equal distribution of all four displacement angles and an equal number of equal and unequal trials. The distribution of stimulus characteristics can be found in Table A.4.

We rendered the figures using the 3D modeling tool Blender [278]. For the 2D condition, we took snapshots in Blender. For the 3D condition, we imported the 3D models into the VR environment. The 3D models could then be displayed, positioned, and rotated there. To compare the 2D and 3D conditions, we used the same combination of base figures and the same rotation angles in each stimulus. The figures' rotation direction and left-right position were varied to reduce memory effects.



(a) Equal pairs. (b) Mirrored unequal pairs. (c) Structural unequal pairs.

Figure A.3.: Examples of our stimulus material with three different types of mental rotation stimuli for 2D (top) and 3D (bottom). Figure sides (left or right) were randomly switched between 2D and 3D to avoid memory effects. The 3D images are screenshots of the VR environment. **Figure 3a.** Equal pairs. **Figure 3b.** Mirrored unequal pairs. **Figure 3c.** Structural unequal pairs.

Characteristic	**Category**	**Number**
Angular disparity	40	9
Angular disparity	80	5
Angular disparity	120	7
Angular disparity	160	7
Stimulus type	equal	14
Stimulus type	mirrored	9
Stimulus type	structural	5

Table A.4.: Characterization of presented stimuli according to their rotation angle (in degree) and their stimulus type.

Apparatus

An HTC Vive Pro Eye and its integrated Tobii eye tracker were used for the VR experiment. The Dual OLED displays inside the HMD provided a combined resolution of 2880 × 1600 pixels, with a refresh rate of 90Hz. The integrated Tobii eye tracker had a refresh rate of 120 Hz and a trackable FOV of 110°, with a self-reported accuracy of 0.5 – 1.1° within a 20° FOV [168]. We ran the VR experiment on a desktop computer using an Intel Core i7 processor with a base frequency of 3.20GHz, 32 GB RAM, and an NVIDIA GeForce GTX 1080 graphic card.

Data collection

While participants used the VR, our data collection pipeline saved stimulus, eye-tracking, and HMD-movement information at each time point, marked with a timestamp. A time point is determined by the VR device's frame rate and the PC's rendering performance. The average frame update rate for all VR runs was 27.31 milliseconds ($SD = 3.36ms$), which translates to 36.61 frames per second. For all experiment runs, the average standard deviation was $6.14ms$. At each frame, we collected eye-tracking data from the Tobii eye tracker, as well as head movement and head rotation. We also noted which stimulus was being presented and if the controllers were being clicked.

We used gaze ray-casting to obtain the 3D gaze points (the location where the eye gaze focuses in the 3D environment). Gaze ray-casting is a method to determine where participants are looking within the scene. For this method, the participant's gaze vector is forwarded as a ray into the environment to see what it intersects with [3], [179]. In our experiment, this gaze intersection was either the virtual screen in the 2D condition or an invisible surface for the 3D condition at the same position.

Data Processing

Data cleaning and pre-processing

After cutting the instructions and tutorial at the beginning of the experiment, we dropped participants with an average tracking ratio below 80% in the raw left and right pupil diameter variables. Since we wanted to compare both conditions (2D and 3D) for each participant, sessions in which only one of the two conditions showed a low tracking ratio also had to be excluded.

The integrated eye tracker already marks erroneous eye detections in the gaze direction variables, which we used to identify missing values. Since blinks are usually not longer than $500ms$ [279], only intervals up to $500ms$ were considered blinks. We needed to detect blinks to correct for artifacts and outliers around blink events [280], [281]. To remove possible blink-induced outliers, we omitted one additional data point around blink intervals, meaning that based on our frame rate, on average, $27ms$ around blinks was missing.

Combined pupil diameter was calculated as the arithmetic mean of the pupil diameter variables for both eyes. A subtractive baseline correction was performed separately for each individual trial. We obtained individual baselines by calculating the median over the 3-second countdown before the stimulus appeared. The values of the combined pupil diameter

during the stimulus intervals were corrected by the baseline measured shortly before. This ensured that potential lighting changes, different background contrasts, or increased fatigue were considered and controlled for [282].

We calculated gaze angular velocity from the experiment data as the change in gaze angle between consecutive points (in degrees per second). The mean distance to the figure was calculated by taking the Euclidean distance between the participant's head location and the midpoint of the stimulus. Additionally, for the 3D condition, we calculated 2D gaze points on an imaginary plane. This plane was set to the same position as the screen in the 2D condition.

Fixation and saccade detection

We applied a combination of a velocity identification threshold (I-VT) and a dispersion identification threshold (I-DT) algorithm for the 2D gaze points [283]. I-VT could be used to detect fixations during stable head movements. However, it was possible to fixate on one spot while rotating one's head around the figure. Because we assumed differences in head movements between the conditions, this would cause artificial differences between conditions. To address this problem of free head movement, we additionally used an I-DT fixation detection algorithm to detect unidentified fixation during periods of head movement.

The I-VT algorithm detected a fixation if the head velocity was $< 7°/s$ and the gaze velocity was $< 30°/s$. We applied the thresholds for each successive pair of data points by dividing the velocity of the gaze or head angles by the time difference between the points. We considered intervals with a duration between $100ms$ and $700ms$ as fixations. We labeled data points as saccades if the gaze velocity was $> 60°/s$ and its duration was below $80ms$. Thresholds for the I-VT algorithm to detect fixation were set conservatively [284]. For the I-DT algorithm, a dispersion threshold of 2° and a minimum duration threshold of $100ms$ were set. To calculate the dispersion, the angle from one data point to another was used, considering the average distance of the participant to the screen or the imaginary surface. Table A.5 shows an overview of the parameters.

Similar threshold parameters for both algorithms have been used in other VR and non-VR studies [37], [283], [284]. The final number of fixations was then formed as a union of both algorithms. We calculated the fixation midpoint for each fixation interval as the centroid point.

	I-DT Fixations	I-VT Fixation	I-VT Saccades
Head velocity (v_h)	-	$v_h < 7°/s$	-
Gaze velocity (v_g)	-	$v_g < 30°/s$	$v_g > 60°/s$
Gaze dispersion (d_g)	$d_g < 2°$	-	-
Duration (Δ)	$\Delta > 100ms$	$100ms < \Delta < 700ms$	$\Delta < 80ms$

Table A.5.: Threshold parameters for detecting fixations and saccades of the velocity and dispersion identification algorithms.

Gaze target information

To calculate features that encode spatial information, for example, on which objects participants fixated, we had to apply further processing steps. This procedure was used to determine whether the fixation location was on or close to one of the figures for each fixation event. If this was the case, the fixation was marked as being on a figure (left or right) and on a specific segment of this figure (inner or outer segment).

Gaze information collected from the VR eye tracker only provides local information about the gaze direction. This means the coordinate system is independent of head movement and head location. The local gaze direction must first be cast into the virtual space by a so-called gaze ray-casting method [3], [195] to get the gaze direction in the virtual space. To find out which object the gaze landed on, the following steps had to be applied. After fixation events are detected, the centers of the fixations hit certain locations in the virtual environment. These locations, also called gaze targets, could either be on the mental rotation figures, close to them, or somewhere else.

Lower accuracy and precision of the HMD produced an offset between the fixation location and the figures. However, we wanted to obtain the most relevant gaze target information. Therefore, fixation locations on a figure, as well as close to a figure, were assigned to that figure. More precisely, for each gaze location, we checked which figure cubes were located close to it. We then checked whether these cubes corresponded to the same segment of the same figure. If the majority of cubes belonged to one segment of one figure, we labeled the fixation location to be on this particular segment. To only assign fixation locations close to the figures, we additionally checked the distance between the fixation locations and the figure centers. If the distance was larger than a radius, we rejected the fixation locations and labeled them as not being on a figure. The radius was obtained by calculating the distance between both figure centers. We calculated the figure centers as the centroid point of all cube midpoints for one figure. Cube midpoints in the 2D condition were based

on manual annotations done by a student assistant with the Computer Vision Annotation Tool `https://github.com/opencv/cvat`. (Retrieved 9/21/2023). To check if all manual annotations were correct, we reconstructed figure plots from the annotation data. Cube midpoints of the 3D figures were collected in the VR environment. An illustration of the process is shown in Figure A.4.

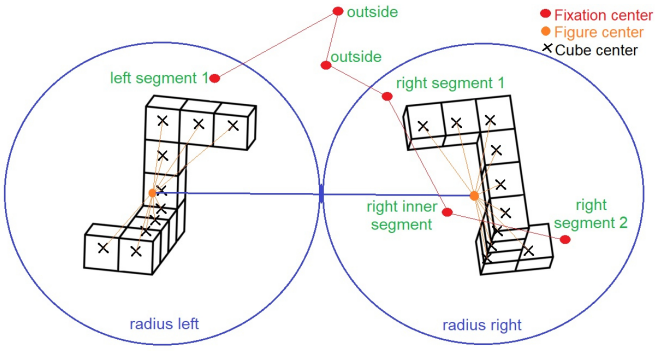


Figure A.4.: A not-true-to-scale illustration of the processing steps involved in finding the closest segments of the figures for each fixation center.

Feature aggregation

Performance measures and condition

Out of the 3024 total presented stimuli (28 stimuli x 54 participants), we needed to remove 46 of these trials due to missing values on at least one feature variable. 2978 trials could be used for the analysis. For each variable, we aggregated the values using the arithmetic mean over all of a person's trials in the 2D and 3D conditions separately.

Reaction time for each trial was calculated using the timestamps in the data. Participants' controller responses were also tracked during the experiment and could be used in combination with a stimulus number to determine a correct or incorrect answer. The experimental data also stored the target variable (2D or 3D).

Eye movement features

Based on the processed experiment data, all eye-movement features were calculated for each stimulus interval separately. For a clearer overview, a description of each feature with the corresponding unit and its calculation is given in Table A.6. We focussed on calculating measures shown to be less affected by sampling errors given a lower sampling frequency

(e.g., fixation duration, fixation rate, and saccade rate) and ignored features like saccade duration [172], [285]. Special attention was paid to the selection of the event detection algorithms to increase reliability by combining two detection algorithms (I-VT and I-DT). We also tried to average out potential outliers by averaging over longer time intervals (Mean fixation duration or mean pupil diameter). To reduce noise and the influence of artifacts on peak pupil diameter, maximum and minimum were only taken within an 80% confidence interval.

Name	**Description**
Mean fixation duration	Average durations in seconds of all fixations within a stimulus interval.
Mean fixation rate	Average over the number of fixations per second.
Mean regressive fixation duration	Average duration in seconds of all fixations after a regressive saccade.
Equal fixation duration between figures	Ratio of the distribution of duration between the figures. Values close to zero indicate that most fixation duration is only on one figure. Values close to one indicate equal fixation duration on both figures (left and right).
Equal fixation duration within figure	Ratio of the distribution of the fixation duration on the whole figure. A value close to zero indicates the most fixation duration on one part of the figures (outer or inner part). A value close to 1 means equal distribution on the outer and inner parts of the figures.
Strategy ratio	Ratio of the number of fixations within the figure divided by the number of saccades between the figure. The number of saccades started as one for the first look at one figure.
Mean saccade velocity	Average over velocities in gaze angle (degree per second) between consecutive time points.
Mean saccade rate	Average over the number of saccades per second.
Mean pupil diameter	Average of all corrected pupil diameter values in millimeters within a stimulus interval.
Peak pupil diameter	Distance between the lowest and highest corrected pupil diameter values in millimeters within an 80% confidence interval of all pupil diameter values within a stimulus interval.
Mean distance to figure	Average Euclidean distance in centimeters from the participants' head to the figure midpoint.
Mean head movement to the sides	Average absolute head movement on the lateral axis in centimeters from the starting position of the participant's head.

Table A.6.: Descriptions of all calculated eye-movement features per stimulus interval.

Data Analysis

Statistical Analysis

The differences between the conditions in some variables were not normally distributed. Thus, we applied a non-parametric, two-tailed, paired Wilcoxon signed-rank test to compare the percentage of correct answers and reaction times between the conditions. We applied the same test for the eye-movement features but corrected the p-values according to Bonferroni's correction. Moreover, we applied a two-tailed, paired t-test for additional verification. The test showed no considerable differences in the p-values for any variables.

Machine learning model

We used a Gradient Boosting Decision Tree (GBDT) classification algorithm to classify the experimental condition since this model had shown high predictive performance in studies with similar data and tasks [177]. Before training the model, we split our data randomly into training and test sets using an 80 to 20 ratio. To increase the reliability of the model performance, we applied a random train-test-split cross-validation with 100 iterations. We trained a GBDT model with eye-movement features at the individual trial level. The model was trained using default hyper-parameters for the Gradient Boosting Classifier from the scikit-learn Python package [286]. We used the 2D or 3D experimental conditions as targets in a binary classification task.

Metrics to evaluate model performance

The within-subject design of the study resulted in almost-balanced sample classes. For the binary classification task (2D and 3D conditions), true positive (TP) cases were correct classifications to the 2D condition, and true negative (TN) cases were correct classifications to the 3D condition (and vice versa for false positives (FP) and false negatives (FN)). The performance metric **accuracy** was calculated as

$$accuracy = \frac{\text{Number of TP} + \text{Number of TN}}{\text{Total Number of Cases}}$$

We report the mean and standard deviation for the accuracy scores over all 100 iterations and for the best-performing model.

Explainability approach

To see how the model uses the measures for prediction, we applied a post-hoc explainability approach using Shapley Additive Explanations (SHAP). Specifically, we used the TreeExplainer algorithm, which computes tractable optimal local explanations and builds on classical game-theoretic Shapley values [178]. Unlike other explainability approaches, which provide information about the global importance of input features, this algorithm computes the local feature importance for each sample. This means we could obtain the importance value for each feature for each classified sample. If a feature exhibited a positive importance value, it drove the model classification towards the positive class and vice versa. The greater the absolute value, the greater its impact on the classification decision. Hence, the overall importance of a feature for classification can be measured by taking the average of the absolute importance values across all samples. Results for local feature importance in the best-performing models are reported in a set of beeswarm plots. The order of the features in the plot represented their overall importance, and each dot displayed the importance and feature value for one sample. Correlated features confound the interpretation of SHAP feature importance for decision tree algorithms. If two features are highly correlated, the algorithm might choose only one feature for prediction and ignore the other completely. Therefore, we checked for multi-collinearity by looking at all measures' pairwise Pearson correlations.

Data availability

The datasets generated and/or analyzed during the current study are available in the osf.io repository, `https://osf.io/vjzmf/?view_only=63de2d2576f04f7cb8059d9669af36c9`

A.1.6. Acknowledgements

Philipp Stark is a doctoral candidate and supported by the LEAD Graduate School and Research Network, which is funded by the Ministry of Science, Research and the Arts of the state of Baden-Württemberg within the sustainability funding framework for projects of the Excellence Initiative II. This research was partly supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC number 2064/1 - Project number 390727645.

A.2. Pupil diameter during counting tasks as potential baseline for virtual reality experiments

A.2.1. Abstract

Pupil diameter is a reliable indicator of mental effort, but it must be baseline corrected to account for its idiosyncratic nature. Established methods for measuring baselines cannot be applied in virtual reality (VR) experiments. To reliably measure a pupil diameter baseline in VR, we propose a short testing environment of visual arithmetic tasks. In an experiment with 66 university students, we analyzed external reliability and internal validity criteria for pupil diameter measures during counting and summation tasks. During the counting task, we found a high retest reliability between stimulus intervals. Acceptable retest reliability was found for task repetition at a second measuring time. Analyzing internal validity, we found that pupil diameter increased with task difficulty comparing both tasks. Further, a linear effect was found between the pupil diameter amplitude and luminance levels. Our findings highlight the potential of counting tasks as a pupil diameter baseline for VR experiments.

A.2.2. Introduction

It has been shown that pupil diameter is a reliable indicator of mental effort, with an increase in pupil diameter associated with an increase in mental effort. This task-evoked pupillary response has been found in arithmetic mathematics [190], reading [191] or memory tasks [128]. Since absolute pupil diameter is idiosyncratic, its values must be corrected by a baseline to allow for between-person comparison [287], [288]. In lab experiments, a baseline can be taken at a resting state at the beginning by fixating on a black screen [115], during stimulus offsets [192] or by calculating a mean value for the entire experiment duration. While these procedures are suited for remote eye trackers, finding a proper baseline for eye tracking in virtual reality (VR) is more complicated. When using a fully immersive VR head-mounted display (HMD), exposing participants to a completely black environment might create fear and discomfort in them [125]. Since emotions also affect pupil diameter [122], this could confound a baseline measurement. If, on the other hand, a baseline is taken at the beginning or for the entirety of the actual experiment, other complications could occur. The aspect of the freely moving interaction with the virtual world by using an HMD, is also a critical aspect when it comes to controlling pupil diameter. Participants could be exposed to different lighting levels depending on where they look in the scene. If

a user decides to look out of the window, directly into a virtual light source for the entire experiment time, then its average pupil diameter will be substantially different from any other user, regardless of idiosyncratic effects. Because pupil diameter is highly influenced by lighting [115], variations in the luminance in the virtual environment could confound a standardized baseline measurement. Moreover, pupil diameter is also influenced by arousal, or mind-wandering [193]. When entering a virtual environment, participants' cognitive states could be different depending on how they perceive the environment and depending on their VR experiences.

To reliably measure a pupil diameter baseline in VR, a controlled testing environment is needed, where participants experience the same level of mental effort, not being negatively influenced or distracted by the immersive surroundings or exposed to different lighting conditions. Therefore, we designed and evaluated a VR environment with two visual arithmetic tasks participants conducted before and after a VR experiment. We propose that this short testing environment could be used to obtain a reliable pupil diameter baseline measurement.

A.2.3. Research Goal and Hypothesis

In our experiment, 66 university students completed two visual arithmetic tasks before and after a VR experience. We measured pupil diameter during simple counting and more complicated summation tasks. We analyzed internal validity and external reliability criteria to evaluate its use as a potential baseline for VR experiments. We argue that during the simple counting task, in which a circle appeared several times at equal intervals, participants maintained a steady, moderate level of concentration and cognitive load without being distracted or letting their minds wander [289]. This should result in consistent, stable average pupil diameter values for each stimulus during the counting period, which led to hypothesis 1.

H1: Average pupil diameter values of participants show high correlations between the stimulus intervals during the counting task.

In a second step, we tested the internal validity of pupil diameter as an indicator of mental load. To see if pupil diameter reacts sensitively to the task's difficulty, pupil diameter should increase with increasing task difficulty. During the summation task, where several circles appear in succession, and their number should be added up, participants should show a higher average pupil diameter than in the counting task. This motivated the following second hypothesis.

H2: The average pupil diameter is significantly larger for the summation task than for the counting task within each participant.

Since pupil diameter is highly sensitive to lighting changes, luminance induced by the visible circles should cause pupillary responses. Due to pupil dilation latency, a drop in pupil diameter should appear 200 – 450 milliseconds after the appearance of the visible circle [115]. In the summation task, different numbers of circles appeared simultaneously. We would expect the drops in pupil diameter to be larger when more luminance is introduced by an increasing number of circles. With pupil diameter amplitude as a measure for drops in pupil diameter, we formulated hypothesis 3.

H3: A significant linear relationship can be found between the number of appearing circles and pupil diameter amplitudes for each stimulus interval during the summation task.

In the summation task, two opposite pupil reactions were to be expected. On the one hand, the overall average pupil size should be larger in the more difficult summation task, because this task requires more cognitive load. On the other hand, locally, when several stimuli appear simultaneously in the task, the pupil should show a stronger response in terms of light-induced contraction. To investigate both effects separately, we formulated the two Hypothesis (H2 and H3). Furthermore, to measure consistency over time, the same participants conducted the counting tasks a second time after a VR experience lasting an average of 11 minutes. To see if participants show the same pupillary response, retest reliability was calculated between both measurement times for all stimulus intervals to test hypothesis 4.

H4: Average pupil diameter values from the same task intervals show high correlations between two measurement times.

A.2.4. Method

Participants and Procedure

We collected data from 66 university students who participated in the experiment. The experiment took place in an experiment lab at the university building. Participants provided written informed consent to participate and received compensation of 10€ after the experiment. Before the VR experiment, a five-point calibration was executed to calibrate the integrated eye tracker. The arithmetic tasks were parts in a more prolonged experiment procedure and were presented before and after a spatial ability experiment also conducted in VR without any

break. Therefore, the same arithmetic tasks were performed at two different measurement times. The time between the first and second testing was, on average, 11.91 minutes (SD=3.65 minutes). In between the two measurement times, participants conducted a mental rotation test, designed as a virtual replication of the original experiment by Shepard and Metzler [230]. An illustration of the VR environment can be seen in Figure 5 in the appendix.

During data processing, 11 participants had to be dropped due to a low tracking ratio of the eye tracker. Details on data processing and exclusion criteria are given in a later section (see subsubsection A.2.4). The remaining 55 participants (33 females and 22 males; $Mean_{age}$ = 24.16, SD_{age} = 7.17) could be used for our analysis. 35 participants used no visual aid during the experiment, 20 used glasses or contact lenses.

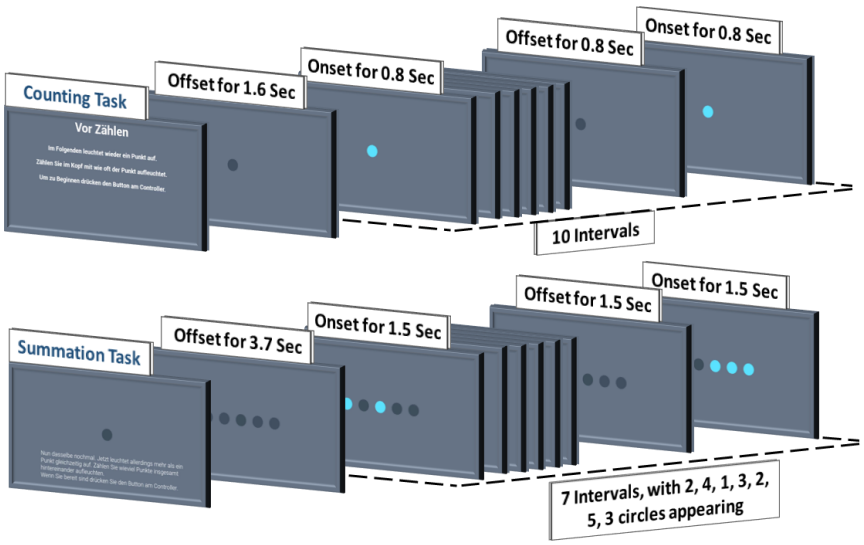


Figure A.5.: Experiment procedure for both tasks. There was always a longer duration before the first stimulus (onset). In the counting task, the number of circles appearing was always one, in the summation task the number of circles varied between one and five.

Design and Apparatus

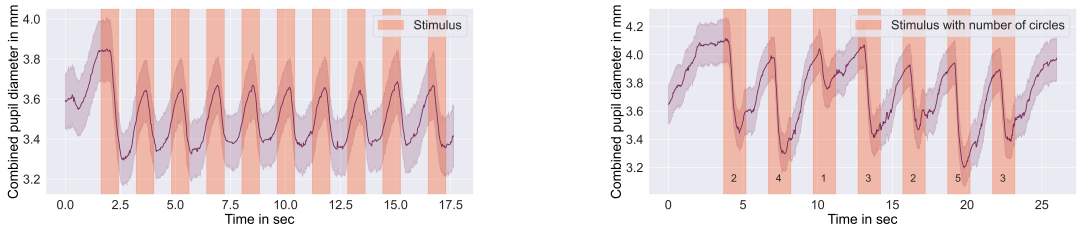
The VR experiment was designed using the Unreal Engine version 4.23 [273]. In both arithmetic tasks (counting and summation), we attached the gray surface to the virtual head of the VR user such that a gray background entirely covered the field of view (FOV). As a result, stimuli that appeared on the surface were always located in the center of participants' FOV

and invariant to head movements. A slightly darker circle marked the center. Before the counting task, participants were instructed that a visible circle would appear several times. They should silently count in their minds how many circles appear and try to move their heads as few as possible. Instructions were written on the gray surface. The task started by clicking on the VR controller, and the first stimulus appeared 1.6 seconds after the click. The stimulus (a blue circle) appeared for 0.8 seconds in the center of the FOV (onset), followed by a 0.8 second interval without a circle (offset). In total, ten circles appeared, one after another. After the counting task, instructions were given for the summation task. Participants were instructed to do the same as in the previous task, but now more than one circle would appear simultaneously on the screen. They should count how many circles appear in total. Since this task was visually more complex, the first stimulus appeared after 3.7 seconds and lasted for 1.5. The onset and offset were set to 1.5 seconds for all seven stimuli. A different number of circles appeared one after the other in the following order: 2, 4, 1, 3, 2, 5, 3 (the complete experiment procedure is displayed in Figure A.5).

An HTC Vive Pro Eye with an integrated Tobii eye tracker was used for the tasks. Inside the HMD, dual-OLED-Displays provided a combined resolution of 2880 × 1600 pixels and a refresh rate of 90Hz. The Tobii eye tracker refresh rate was 120 Hz with a trackable FOV of 110° [168]. Since the pipeline for eye-tracking data collection was integrated into the VR environment, the sampling rate was limited by the VR's update rate, which amounted to an average sampling rate of 45Hz.

Data Processing

From the integrated eye tracker, we obtained pupil diameter values for the left and right eye in millimeters (mm) on average every 25 milliseconds. Raw left and right pupil diameter values were processed according to proposed guidelines for pupillometry [280]. Signal artifacts during blinks and outliers were removed using a blink reconstruction function [290]. This *blinkreconstruct* function was evaluated for high-frequency eye trackers [291]. Because our VR eye tracker had a lower sampling rate, we changed two default parameters in the function. We lowered the pupil-velocity threshold *vt start* from 10 to 5, which resulted in more easily triggered blinks. We also lowered the *gap margin* number from 20 to 5, which determines how many data points around the missing data are not reconstructed. For all other arguments in the function, we used default parameters. After processing the pupil diameter values, we removed participants if their pupil diameter tracking ratio for the left or right eye was lower than 80%. Additionally, we checked for the mean difference in pupil diameter between the



(a) Ten stimulus intervals during the counting task.

(b) Seven stimulus intervals during the summation task.

Figure A.6.: Average pupil diameter (purple line) and standard deviation (transparent purple area) in millimeters for both tasks at the first measurement time. The orange bars show the stimulus onsets. For the summation task, the number of appearing circles per stimulus interval is written at the bottom of the stimulus bar.

left and right eyes to detect potential inaccuracies in the eye tracker and set an empirically motivated threshold of $0.9mm$. For one participant, the mean difference was greater than $0.9mm$, so we removed this participant from the sample. After that, a combined pupil diameter variable was calculated as the average over both eyes. Applying the exclusion criteria, 55 participants could be used for analysis.

Measurements and Analysis

We split each task into equally sized stimulus intervals. To account for pupil dilation latency, we did not set interval boundaries based on the margins of stimulus appearance (onset). We started and ended a stimulus interval in the middle of each stimulus offset when no stimulus was present. So each stimulus interval was one-half of the offset before the onset and onset and one-half of the offset afterward. We calculated the average pupil diameter per participant for each stimulus interval as the mean of all pupil diameter values. We also calculated pupil diameter amplitude per participant as the difference between the minimum and maximum pupil diameter value within each stimulus interval. Artifact outliers can confound minimum and maximum values. Therefore, we calculated the median for the five largest and smallest values and compared them to the maximum and minimum values. As descriptive statistics, we report an overall mean value for all participants per task and standard deviations between stimulus intervals and between participants. For better visualization, we created Figure A.6a and Figure A.6b, showing participants' average pupil diameter and standard deviation at each time point for the counting and the summation task at the first measurement time. For

the second measurement time, values are visualized in the appendix in Figure 4a and Figure 4b.

To compare pupil diameter measures for stimulus intervals within and between the tasks, we computed Pearson's correlation with Bonferroni-Holmes corrected p-values. Correlation coefficients were used to evaluate external reliability. We also calculated the total mean pupil diameter across all stimulus intervals within each task. To report differences in average pupil diameters between both tasks, we applied a Wilcoxen-rank test (because values were not normally distributed). To compare the different luminance levels in the summation task, we evaluated a simple linear regression model with the number of appearing circles as the independent variable and the pupil diameter amplitude as the dependent variable. The results of the difference measures and the regression were evaluated as criteria for internal validity.

A.2.5. Results

For the *counting task before the VR experience*, the mean pupil diameter over all stimuli intervals and participants was $3.501mm$. The standard deviation between stimuli intervals was $0.051mm$ and between participants $0.511mm$. For the *summation task before the VR experience*, the mean pupil diameter over all stimuli intervals and participants was $3.687mm$. The standard deviation between stimuli intervals was $0.091mm$ and between participants $0.529mm$. The mean pupil diameter amplitude in the summation task was $0.740mm$, with a standard deviation between stimuli intervals of $0.118mm$ and between participants of $0.188mm$.

For the *counting task at the second measurement time* after the VR experience, the mean pupil diameter over all stimuli intervals and participants was $3.364mm$. The standard deviation between stimuli intervals was $0.040mm$ and between participants $0.528mm$. For the *summation task after the VR experience*, the mean pupil diameter over all stimuli intervals and participants was $3.521mm$. The standard deviation between stimuli intervals was $0.070mm$ and between participants $0.535mm$. The mean pupil diameter amplitude in the summation task was $0.721mm$, with a standard deviation between stimuli intervals of $0.113mm$ and between participants of $0.184mm$. Boxplots showing the central tendencies of pupil diameter values between participants, tasks and measurement times can be seen in Figure A.7.

Testing retest reliability for the first measurement time (**H1**), correlation coefficients of the average pupil diameters between the stimulus intervals in the counting task ranged from $r = 0.855$ to $r = 0.978$. All correlations were significant (with corrected p-values, $p < 0.001$)

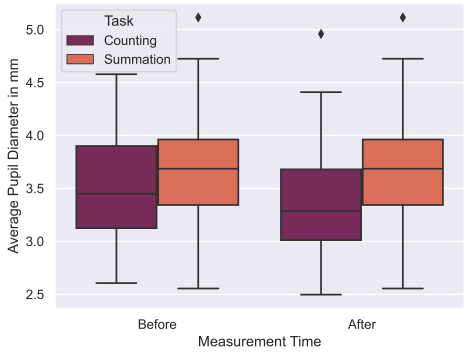


Figure A.7.: Boxplots of the average pupil diameters for all participants for both tasks (counting and summation task) and at both measurement times (before and after the VR experience).

and can be found in the lower triangle in Table A.7. Similar correlations were obtained for the second measurement time and can be found in the upper triangle in the same table (also see Table A.7).

Testing for differences between the counting and the summation task (**H2**), the Wilcoxon-rank test indicated that the mean pupil diameter was statistically significantly higher in the summation task ($Z = 1490$, $p < 0.001$). Similar differences could be found for the second measurement time ($Z = 1409$, $p < 0.001$).

We used a simple linear regression to test if the number of appearing circles (NAC) significantly predicted pupil diameter amplitude (**H3**). The fitted regression model was: $0.5362 + 0.0713 * (NAC)$. The overall regression was statistically significant ($R^2 = 0.113$, $F(1, 383) = 48.66$, $p < 0.001$). The NAC significantly predicted pupil diameter amplitude ($\beta = 0.0713$, $p < 0.001$).

Lastly, we compared the average pupil diameter values for the counting task intervals between the two measurement times (**H4**). We found statistically significant correlations for all ten counting intervals ranging from $r = 0.728$ to $r = 0.837$ (with corrected p-values, $p < 0.001$). All results can be seen in Table A.8.

A.2.6. Discussion

To overcome difficulties in measuring a reliable pupil diameter baseline for VR experiments, we proposed using an arithmetic counting task conducted at the beginning of a VR exper-

Table A.7.: Correlation table of participants' average pupil diameter values between stimulus intervals (SI) during the counting task. The lower triangle reports Pearson's correlation coefficients during the first measurement time before the VR experience. The upper triangle reports Pearson's correlation coefficients during the second measurement time after the VR experience. *** indicates Bonferoni-Holmes corrected p-values with $p < 0.001$.

	SI 1	SI 2	SI 3	SI 4	SI 5	SI 6	SI 7	SI 8	SI 9	SI 10
SI 1	-	0.964^{***}	0.954^{***}	0.940^{***}	0.930^{***}	0.925^{***}	0.921^{***}	0.946^{***}	0.926^{***}	0.915^{***}
SI 2	0.923^{***}	-	0.977^{***}	0.951^{***}	0.927^{***}	0.925^{***}	0.938^{***}	0.944^{***}	0.923^{***}	0.911^{***}
SI 3	0.921^{***}	0.978^{***}	-	0.965^{***}	0.929^{***}	0.920^{***}	0.925^{***}	0.929^{***}	0.915^{***}	0.904^{***}
SI 4	0.916^{***}	0.950^{***}	0.967^{***}	-	0.965^{***}	0.923^{***}	0.903^{***}	0.924^{***}	0.910^{***}	0.889^{***}
SI 5	0.876^{***}	0.907^{***}	0.923^{***}	0.962^{***}	-	0.963^{***}	0.911^{***}	0.916^{***}	0.894^{***}	0.882^{***}
SI 6	0.878^{***}	0.937^{***}	0.932^{***}	0.941^{***}	0.945^{***}	-	0.943^{***}	0.938^{***}	0.903^{***}	0.879^{***}
SI 7	0.884^{***}	0.921^{***}	0.917^{***}	0.923^{***}	0.915^{***}	0.971^{***}	-	0.970^{***}	0.928^{***}	0.902^{***}
SI 8	0.870^{***}	0.909^{***}	0.910^{***}	0.918^{***}	0.916^{***}	0.956^{***}	0.964^{***}	-	0.967^{***}	0.938^{***}
SI 9	0.868^{***}	0.906^{***}	0.908^{***}	0.920^{***}	0.936^{***}	0.947^{***}	0.942^{***}	0.968^{***}	-	0.966^{***}
SI 10	0.855^{***}	0.904^{***}	0.902^{***}	0.918^{***}	0.924^{***}	0.93^{***}	0.938^{***}	0.941^{***}	0.977^{***}	-

Table A.8.: Correlation table comparing both measurement times, separately for each stimulus interval (SI). The average pupil diameter values of the participants during the counting task were correlated. We report Pearson's correlation coefficients, where *** indicates Bonferoni-Holmes corrected p-values with $p < 0.001$.

SI 1	SI 2	SI 3	SI 4	SI 5	SI 6	SI 7	SI 8	SI 9	SI 10
0.837^{***}	0.814^{***}	0.825^{***}	0.775^{***}	0.754^{***}	0.745^{***}	0.746^{***}	0.736^{***}	0.728^{***}	0.736^{***}

iment. Our study analyzed pupil diameter values from 55 university students during two arithmetic tasks (counting and summation). We evaluated internal validity and external reliability criteria. For the counting task, we analyzed the retest reliability of average pupil diameters between the stimulus intervals (H1). We also investigated if the increased task difficulty during the summation task led to a higher average pupil diameter (H2) and if luminance levels could predict pupil diameter drops during task solving (H3). Lastly, we tested retest reliability by comparing participants' pupil diameters between the counting task at the beginning with the same task performed at the end of a VR experiment (H4).

All Persons' correlation coefficients for stimulus intervals during the counting task at the beginning showed good to excellent reliability ($r > 0.8$ or $r > 0.9$). Therefore, we did not reject **Hypothesis 1**. For the first stimulus interval, correlation coefficients were slightly lower compared to the other intervals, which might have been caused by participants' unfamiliarity with the task. For the counting task at the second measurement time, the first interval showed

correlations greater than $r = 0.9$, which supports this assumption (see Table A.7). Because we found a significant increase in average pupil diameter for the summation task, we also did not reject **Hypothesis 2**. Even though pupil diameter amplitudes were larger in the summation task, we could still measure a significant increase in average pupil diameter. We also found a significant linear relationship between the number of circles and the pupil diameter amplitude, which explained 11.3% of variance in the data. So we could argue that drops in pupil diameter found in both tasks resulted from the induced luminance. Therefore, we did not reject **Hypothesis 3**. Additionally, we found that the retest reliability comparing both measurement times was only acceptable ($r < 0.7$) for most intervals. Therefore, we rejected **Hypothesis 4**. Descriptive statistics showed a decrease in average pupil diameter for the second measurement time. Further investigation might be necessary to determine whether this difference was caused by a decreasing inaccuracy of the eye tracker or by fatigue or drowsiness [115]. A second eye tracker calibration could be performed before the second measurement time to distinguish this effect.

One limitation of the study was that we did not control whether participants performed the tasks correctly or not. Weighing different design aspects, we implemented the counting tasks to confound the main experiment as little as possible. We wanted participants to stay relaxed and avoid any feeling of examination or performance testing that could influence later experiments in VR. Moreover, one must ask whether people who unintentionally count incorrectly do not feel the same mental effort because they are unaware of their incorrect result. In a sense, performance testing for different levels of arithmetic competence might not be a good indicator of the perceived mental effort. By checking participants' performance, we might only detect those not performing the experiment seriously. To gain more insights here, a multi-modal design could be used in future experiments. Measuring levels of mental effort using other indicators (e.g., from EEG data) could validate our results. Another limitation regarding the experiment design was that we did not randomize the order of the two tasks. Possible crossover effects could be detected in the future with a counterbalancing design.

Despite these limitations, we demonstrated that participants showed consistent, stable average pupil diameter values over time during the visual counting task. The increase in pupil diameter in the summation task suggested that participants experienced less mental effort during counting. We could demonstrate an overall increase in mental effort with increased task difficulty and the effect of a light-inducted pupil contraction. Both results helped to explain and validate the pupil diameter patterns during the counting task and indicated that this task is reliable and valid for baseline use.

A.2.7. Conclusion

We evaluated that a reliable baseline for measuring pupil diameter in VR can be obtained using an experiment environment for visual arithmetic tasks. This testing procedure takes little time, is quick to implement, and is minimally affected by factors such as head movements, varying luminance levels, and mental states. The baseline can be calculated by averaging the mean intervals during the counting task. Further approaches could incorporate differences in pupil diameter between tasks of varying difficulty to account for changes in pupil diameter. A subtractive or divisive baseline correction can be applied to control for idiosyncratic effects. Controlling for the effect of luminance on pupil size, although a separate issue, can be applied independently in addition to the idiosyncratic standardization proposed in this study.

A.2.8. Acknowledgments

Philipp Stark is a doctoral candidate and supported by the LEAD Graduate School and Research Network, which is funded by the Ministry of Science, Research, and the Arts of the state of Baden-Württemberg within the framework of the sustainability funding for the projects of the Excellence Initiative II.

B. Visual Attention in a Virtual Classroom

The following publications are enclosed in this chapter:

[3] E. Bozkir, **P. Stark**, H. Gao, L. Hasenbein, J.-U. Hahn, E. Kasneci, and R. Göllner, "Exploiting object-of-interest information to understand attention in VR classrooms", in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, Mar. 2021, pp. 597–605. DOI: 10.1109/VR50410.2021.00085

[4] **P. Stark**, A. Jung, J.-U. Hahn, E. Kasneci, and R. Göllner, "Using gaze transition entropy to detect classroom discourse in a virtual reality classroom", in *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications (ETRA '24)*, Glasgow, UK: ACM, 2024. DOI: 10.1145/3649902.3653335

In [3] the first and second author contributed equally. Publications are included with format modifications. Definitive versions are available via digital object identifiers at the relevant venues.

B.1. Exploiting object-of-interest information to understand attention in VR classrooms

B.1.1. Abstract

Recent developments in computer graphics and hardware technology enable easy access to virtual reality headsets along with integrated eye trackers, leading to mass usage of such devices. The immersive experience provided by virtual reality and the possibility to control environmental factors in virtual setups may soon help to create realistic digital alternatives to conventional classrooms. The importance of such settings has become especially evident during the COVID-19 pandemic, forcing many schools and universities to provide the digital teaching. Researchers foresee that such transformations will continue in the future with virtual worlds becoming an integral part of education. Until now, however, students' behaviors in immersive virtual environments have not been investigated in depth. In this work, we study students' attention by exploiting object-of-interests using eye tracking in different classroom manipulations. More specifically, we varied sitting positions of students, visualization styles of virtual avatars, and hand-raising percentages of peer-learners. Our empirical evidence shows that such manipulations play an important role in students' attention towards virtual peer-learners, instructors, and lecture material. This research may contribute to understanding of how visual attention relates to social dynamics in the virtual classroom, including significant considerations for the design of virtual learning spaces.

B.1.2. Introduction

Everyday use of head-mounted displays (HMDs) is increasing as virtual reality (VR) technology and virtual environments are already being used in various domains such as gaming and entertainment. In addition, some of the consumer-grade HMDs are coming to market with integrated eye trackers that may help to assess human attention during immersion and allow for more interactive virtual environments. It is likely that, in the near future, such tools will become widely used mobile devices similar to today's mobile phones or smart watches. To this end, not only should researchers strive to improve the capabilities of these devices, but scrutiny should also be given to understanding human behavior and attention while using such technology.

Measures of eye movements obtained through eye-tracking are effective indicators of human states and visual behavior to some extent; however, they are dependent on application

or task [292]. Analyzing and modeling human attention using this data in a specific domain may not be transferable to other domains. Thus, when assessing human attention in digital environments, or more particularly in VR for the application in educational technology, specific domain knowledge and configurations should be considered. There is already some history of training and teaching in digital or virtual setups [293], [294]. Today, due to the COVID-19 pandemic, virtual or digital education has become more popular and even a necessity in many cases. Currently, many schools and universities are carrying out their teaching responsibilities remotely via platforms such as Zoom[1] or Webex[2]. Such platforms lack the possibility of instructor-student interaction beyond audio and video features and encounter privacy concerns if videos are recorded and stored during classes. VR setups offer the immersion, interaction, and privacy preservation that current remote learning platforms lack. In addition, as VR allows users to easily control the environmental settings, it is possible to evaluate different classroom manipulations and subsequent effects on human behavior, a step that is exponentially more difficult in real world classrooms.

In this work, we exploit object-of-interest information by using eye-gaze and three main sets of objects in immersive VR. We focus on virtual peer-learners, virtual instructor, and screen to understand visual attention through the design of a virtual classroom and a lecture about computational thinking. We choose these objects-of-interests since they are of particular interest with regard to attention towards social dynamics and learning. Our study has three different design factors: Different sitting positions of participating students, different visualization styles of virtual avatars including an instructor and peer-learners, and different hand-raising behaviors of virtual peer-learners. Different sitting positions include seating participating students in the front or back of the virtual classroom. In addition, different visualization styles of avatars consists of two conditions that are cartoon- and realistic-styled avatars. Lastly, different hand-raising behaviors include 20%, 35%, 65%, and 80% of the peer-learners raising their hands to answer questions during the lecture. To the best of our knowledge, this is the first work that assesses students' attention by using object-of-interest information in an immersive VR classroom through the manipulation of sitting positions of students, visualization styles of peer-learners and instructor, and hand-raising behaviors of peer-learners collectively. Such manipulations may be important indicators of students' visual attention towards lecture contents and social dynamics in the classroom and should be taken into consideration when designing VR classrooms.

[1] https://www.zoom.us/
[2] https://www.webex.com/

B.1.3. Related Work

Since our work benefits from VR in education and in eye tracking research, we discuss the state-of-the-art along these two lines. Various studies using VR in education settings assess the mechanisms of attention or social dynamics by using pre- or post-tests or by relying on head movement behavior as a proxy for gaze. Using eye tracking in addition to such information presents the possibility of a deeper understanding of visual and situational attention during immersive experiences.

Virtual Reality in Education and Classrooms

VR offers great promise for supporting teaching and learning procedures, especially when digital learning, physical inabilities, ethical concerns, and situational limitations are considered. An extensive review of immersive VR in education and its pedagogical foundations are discussed in [294] and [28], respectively. We focus on research on VR in education and immersive VR classrooms in this section.

The effectiveness of learning in virtual and augmented reality (VR/AR) compared to tablet-based applications and the impact of VR-based systems on students' achievements are studied in [295] and [296], respectively, and these works indicate several advantages of VR-based conditions. In addition, it has been found that students' motivation increases when VR is used as a teaching tool in art history [297] and social studies [298]. VR not only supports the effectiveness of learning, but also can improve instructor teaching skills [299].

Apart from VR applications in teaching and learning, the design and degree of realism in VR classrooms have also been studied. Presence of a virtual instructor was found to increase the engagement and progress of users [300]. Furthermore, the processes of synthesizing virtual peer-learners by using previous learner comments [301] and designing VR classrooms by replicating real conditions [302] which may affect learning are considered.

Several works focused on understanding visual attention and behavior in immersive VR classrooms. Bailenson et al. [91] and Blume et al. [66] studied learning outcomes according to sitting positions and offer compelling evidence that students seated in the front have better learning outcomes. Few studies, however, took head movements into consideration [67], [76], [93], [96] in such setups. In [93], the immersive VR classroom was used as a tool to study attention measures for attention deficit/hyperactivity disorder (ADHD), whereas in [96] reliability of virtual reality and attention was studied with continuous performance task (CPT) for clinical research. Social interaction using head movements was studied in [67] with users' head movements found to shift between the interaction partner and target. Some

studies argued for eye tracking measurements, especially in clinical research for diagnosis or attention related tasks [94], [97]. However, none of the previous works have focused on social interactions and dynamics in the immersive VR classroom in an everyday setting by using object-of-interest information and eye movements.

Eye Tracking in Virtual Reality

Eye tracking and gaze estimation are considered challenging tasks in a real world setting because it is difficult to control factors such as occlusions or illumination changes [303], [304]. However, in most of the VR setups, eye trackers are located inside of HMDs. This creates not only a more controlled and reliable environment for eye tracking, but also provides a unique opportunity to analyze and process human visual behavior during the VR experience.

Eye tracking has been used in many applications and shown to be helpful for various tasks in VR such as guiding attention in panoramic videos using central and peripheral cues [305], predicting motion sickness by using 3D Convolutional Neural Networks [306], synthesizing personalized training programs to improve skills [307], foveated rendering using saccadic eye movements and eye-dominance [308], [309], evaluation and diagnoses of diseases such as Parkinson's disease [310], re-directed walking using blinking behavior [311], or continuous authentication using eye movements [312]. While these works have used either the eye tracking or gaze data to derive more meaningful information for related tasks, assessing visual attention via eyes and gaze-based interaction is more relevant for classroom setups in particular. Bozkir et al. [313] assessed visual attention using gaze guidance and pupil dilations in a time-critical situation, whereas Khamis et al. [314] discussed gaze-based interaction using smooth pursuit eye movements in VR. In addition, Sidenmark and Lundström [315] analyzed eye fixations on interacted objects during hand interaction in VR and found that interaction with stationary objects may be favorable. Aforementioned works indicate that eye movements can be used reliably in VR setups. Moreover, considering that the majority of objects in a classroom are stationary or have limited spatial movement, visual attention extracted from such data may provide valuable insight into human behavior. While exploiting objects-of-interests could be considered as a primitive task, it forms the foundation of more complex tasks necessary to understand visual attention.

B.1.4. Methodology

The main focus of this work is to investigate object-of-interest information in different manipulations of an immersive VR classroom. We focus on three objects that may be considered

(a) Overall virtual classroom design.

(b) Hand-raising cartoon-styled peer-learners from back.

(c) Realistic-styled peer-learners.

(d) Hand-raising cartoon-styled peer-learners.

Figure B.1.: Views from the virtual classroom.

as the most important objects in the current setup, namely peer-learners, instructor, and screen.

Participants

381 volunteer sixth-grade students (179 female and 202 male) between 10 to 13 years old ($M = 11.5$, $SD = 0.6$) were recruited for the experiment. In this age group, students are able to use an HMD, but do not have much experience with VR. They also had no background knowledge about the lecture content. Data from 101 participants were removed due to hardware related problems, incorrect calibration, low eye tracking ratio (lower than 90%), and synchronization issues. The average number of participants per condition was 17.5 ($SD = 5.2$). Finally, we used the data of 280 participants (140 female and 140 male) with the aforementioned average age and standard deviation. For each condition group separately, participants' gender was also equally distributed ($M = 0.58, SD = 0.08$). The study was approved by the ethics committee of the University of Tübingen prior to the experiments. Participants and their parents or legal guardians provided written informed consent in

advance.

Apparatus

For the experiments, HTC Vive Pro Eye devices with integrated Tobii eye trackers were used. The HTC Vive Pro Eye has a refresh rate of 90 Hz and field of view of 110°. The integrated eye tracker has 120 Hz sampling rate. The screen resolution per eye was set to 1440 × 1600. Unreal Game Engine v4.23.1[3] was used to render the virtual classroom.

Experimental Design

The virtual classroom consists of 4 rows of desks organized in 2 columns. Next to each desk, chairs are located to let virtual peer-learners sit. There are 24 virtual peer-learners in the environment and all of them sit on chairs during the entirety of the lecture. Some of the chairs are kept empty so as not to overcrowd the virtual classroom. In addition, the virtual classroom includes other objects, which exist in real classrooms such as board, screen, cupboard, clock, and windows. The lecture content is visualized on the white screen. Additionally, the virtual instructor walks around the podium, replicating behavior similar to that of a real instructor. Figure B.1 (a), (b), (c), and (d) show the overall design, hand-raising peer-learners, realistic-styled peer-learners, and cartoon-styled peer-learners, respectively.

The content of the virtual lecture is about computational thinking [316] and the lecture takes ≈ 15 minutes in total, including 4 phases. These four phases are grouped as "Introduction to the topic", "Knowledge input", "Exercises", and "Summary" and take ≈ 3, ≈ 4.5, ≈ 5.5, and ≈ 1.5 minutes, respectively. The topic of the virtual lecture is visible on the board as "Understanding how computers think". The first phase starts with the virtual instructor entering the classroom. After staying for a while, the instructor leaves the classroom for about 20 seconds. During this time, participants have the opportunity to explore the classroom, look around, and acclimate themselves with the virtual environment. During the initial phase of the lecture, the instructor asks five questions, and some of the virtual peer-learners raise their hands to interact. In the second phase, the instructor describes two terms, "sequence" and "loop", and shows these terms on the white screen. After the descriptions, the instructor asks four questions about each term and some of the peer-learners raise their hands to answer them. In the third phase, the instructor assigns two exercises and allows students some time to think about them. Later, choices for each exercise are provided by the instructor and, this time, peer-learners raise their hands to vote on the correct answer out of

[3] https://www.unrealengine.com/

the presented options. In the fourth phase, the instructor summarizes the lecture without asking any questions, which means that peer-learners do not raise their hands. In addition, no hand-raise is expected from the participants as hand poses are not measured during the experiments.

Our study is conceptualized in a between-subjects design. We evaluated three design factors, namely sitting positions of the participants, visualization styles of virtual avatars, and hand-raising percentages of virtual peer-learners. Participants were seated either in the front or back rows, which means that the participants seated in the front had one row in front of them, whereas participants seated in the back had three rows between them and the screen. Both conditions were aligned in the aisle side of the desks that were on the right side of the classroom. This manipulation can give insights about students' attention during a lecture, when they have either the overview over whole class and see most of their virtual peer-learners or when they are positioned closer to instructor and screen the lecture is presented on. Participants encountered either cartoon- or realistic-styled virtual avatars in the environment, including the virtual instructor and peer-learners. The cartoon-styled avatars have larger heads and tinier arms and legs as compared to the realistic-styled avatars. Since the animation and design of more realistic looking avatars is time and cost expensive, it should be interesting to investigate the impact of such manipulation. In addition, various hand-raising percentages of virtual peer-learners consist of four levels, namely 20%, 35%, 65%, and 80%. This means that when a question is asked during the lecture by the virtual instructor, a corresponding percentage of virtual peer-learners raise their hands to answer the question. The last two manipulations are of particular interest, regarding the question how social avatars should be designed in a virtual classroom and how they are perceived by students. Under which condition do students use social information and how does visualization and certain behaviour influence students attention. This helps to simulate and evaluate social dynamics and engagement during the virtual lecture using visual attention. In total, our 2 (factor 1) $\times$2 (factor 2) $\times$4 (factor 3) between-subjects design leads to 16 treatment groups.

Procedure

In the beginning of the experiment, the assistants introduced the experiment and its process to the participants. Participants had the opportunity to familiarize themselves with the hardware and the VR environment. Afterwards, the actual experiment and data collection began. Firstly, the eye tracker was calibrated. Then, the experiment was started with assistants

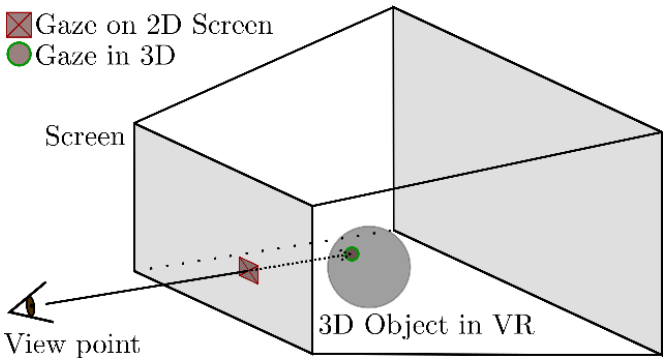


Figure B.2.: Ray-casting procedure to obtain 3D gazed object.

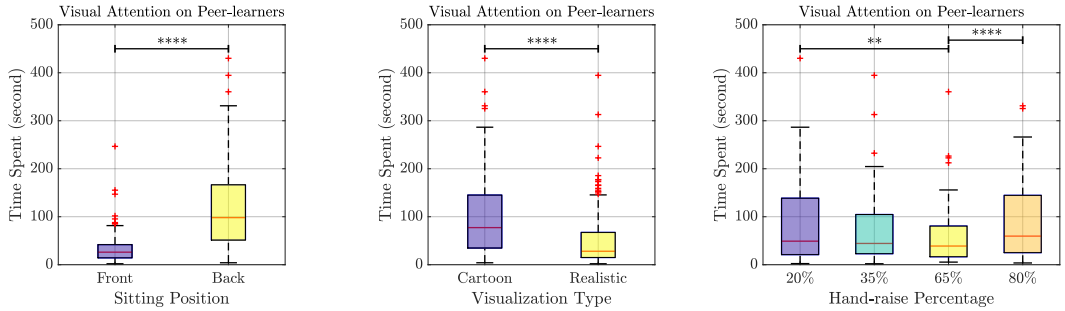
pressing a start button. At the end of the virtual lecture, the participants were told to take the HMD off by a message which was displayed in the virtual environment. Virtual lectures were carried out without any breaks. After watching the virtual lecture, participants filled out questionnaires about their perceived realism and experienced presence which were conceptualized for the VR classroom according to [79], [317].

Each session took ≈ 45 minutes in total. The experiments were carried out in groups of ten participants who were randomly allocated to one of the 16 treatment groups by using a random number generator to ensure the random distribution of conditions within groups. To maintain natural behavior, participants selected the physical seat in the experiment room freely without being informed about experimental conditions. Although research assistants helped with technical issues regarding the use of the HMD, participants were blinded to the true purpose and design of the study, as it was solely introduced as a learning experience.

Data Processing and Measurements

During the experiments, head location and pose, gaze, and eye related data along with experimental condition were collected. Head movements are particularly helpful for mapping eye-gaze in the virtual environment. These were saved in data sheets for each participant using anonymous identifiers which ensured the privacy of the participants.

As gaze data reported by the eye tracker can be affected negatively by blinks or noisy sensor measurements, we applied a linear interpolation on the gaze vectors to clean the data. Afterwards, using head pose and interpolated gaze data, we applied ray-casting [318] to map the gaze into the 3D virtual environment. The objects in the 3D environment are surrounded by dedicated colliders; therefore, we were able to calculate 3D gaze points and gazed objects



(a) Comparison between sitting positions.

(b) Comparison between visualization types.

(c) Comparison between hand-raising behaviors.

Figure B.3.: Attention towards virtual peer-learners for different classroom manipulation configurations. $*$, $***$, and $****$ correspond to the significance levels of $p < .05$, $p < .001$, and $p < .0001$, respectively.

using the procedure visualized in Figure B.2.

However, gazed objects may not directly represent visual attention as participants can gaze on some objects unconsciously for a very short time. To overcome this issue, we set an attention threshold of 200 ms, meaning that we count the objects as object-of-interest if participants stay with their gaze on the objects for at least the amount of the attention threshold. As we assume that both fixations and saccades can occur during attending one object, the selected threshold is larger than classical fixation thresholds applied in eye tracking literature for both conventional [283] or VR eye tracking [284] setups. While we also experimented with various threshold values, our results show similar trends across different thresholds.

In addition to the data related to visual attention, self-reported perceived realism and experienced presence were obtained at the end of the experiments with 4-point Likert scales ranging from 1 ("completely disagree") to 4 ("completely agree") with 6 (e.g., "I felt like the teacher and the classmates could be real people") and 9 (e.g., "During the virtual lecture, I almost forgot that I was wearing the VR glasses") items, respectively.

In this study, we focused on three main objects in the virtual classroom, namely peer-learners, virtual instructor, and screen, when we extracted object-of-interest information. We decided that these objects may have a significant impact on social dynamics in the classrooms and for overall course of lecture. In our analyses, the attention time on each peer-learner is aggregated and the object of "peer-learners" represents the aggregated object and related attention. In addition, in our classroom setup there is one board and one white screen behind the instructor as depicted in Figure B.1 (a). The lecture content is provided on

the white screen only; therefore, in our analysis we refer to the white screen when mentioning screen object.

Research Hypotheses

Our hypotheses correspond to the experimental factors of sitting positions, avatar visualization styles, and various hand-raise percentages of virtual peer-learners, respectively. Furthermore, since we analyze behaviors towards three different objects in the virtual classroom, namely peer-learners, instructor, and screen, for simplicity we call attention to attending these objects-of-interests for the rest of the paper.

Visual Attention in Different Sitting Positions (H1)

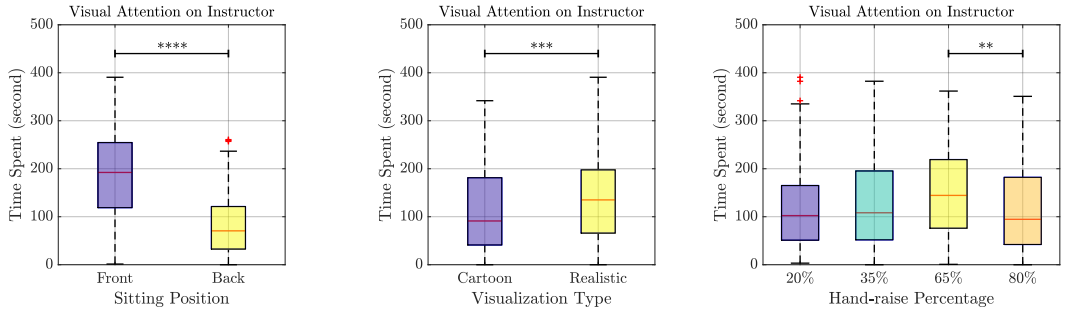
We expect that participants seated in the front condition have less attention on peer-learners, naturally because they do not have as many peer-learners sitting in front of them as opposed to the participants sitting in the back. In addition, the participants that are located in the front are closer to the virtual instructor and the screen that visualizes lecture content. Due to the proximity and having fewer moving and occluding objects in their field of view (FOV), we hypothesize that these participants have more attention time on both virtual instructor and screen than the participants sit in the back.

Visual Attention in Different Visualization Styles of Virtual Avatars (H2)

We hypothesize that attention time on peer-learners in the cartoon-styled visualization is longer than in the realistic-styled visualization as cartoon-styled peer-learners are more exciting for participants when ages of our interest group are taken into consideration. In addition, we assume that participants look at the realistic-styled instructor for longer than at cartoon-styled instructor as participants may consider the realistically rendered instructor more credible in a learning environment. Lastly, we do not expect any differences in terms of attention towards virtual screen that lecture content is visualized, as the visualization style of the screen does not change.

Visual Attention in Different Hand-raising Behaviors of Peer-learners (H3)

We hypothesize that attention time on peer-learners increases with a higher number of virtual peer-learners raising their hands when questions are asked, as this would create a visually more dynamic classroom. Additionally, we expect that if fewer virtual peer-learners raise their hands, this will lead participants to keep their attention either on the instructor



(a) Comparison between sitting positions.

(b) Comparison between visualization types.

(c) Comparison between hand-raising behaviors.

Figure B.4.: Attention towards virtual instructor for different classroom manipulation configurations. $*$, $***$, and $****$ correspond to the significance levels of $p < .05$, $p < .001$, and $p < .0001$, respectively.

or the lecture screen due to having less amount of visual distractors when questions are provided by the virtual instructor.

B.1.5. Results

In this section, we analyze the total amount of time spent on each object-of-interest (OOI), which we call visual attention, between different conditions. For each OOI, we applied a 3-way full factorial ANOVA for statistical comparison using alpha level of 0.05. For non-parametric analysis, we transformed the data using the aligned rank transform (ART) [319] before applying ANOVAs. For the pairwise comparisons, we used Tukey-Kramer post-hoc test as the sample sizes were not equal. While the main focus of this work is to assess visual attention using OOI information, here we report experienced presence and perceived realism questionnaires to support our main results. We obtained mean values of 2.91 for experienced presence and perceived realism with $SD = 0.55$ and $SD = 0.57$, respectively, without any significant differences between conditions.

Visual Attention on Peer-learners

Total time spent on peer-learners for different sitting positions, avatar visualization styles, and various hand-raising behaviors are depicted in Figure B.3 (a), (b), and (c), respectively. Total time spent on peer-learners is significantly longer in the back seated condition ($M = 115.07$ sec, $SD = 85.28$ sec) than it is in the front seated condition ($M = 33.59$ sec, $SD = 32.45$ sec) with ($F(1,264) = 156.23$, $p < .0001$, $\eta^2 = .36$).

Attention towards peer-learners as different visualization styled avatars differs significantly. Cartoon-styled peer-learners (M = 98.67 sec, SD = 82.79 sec) drew significantly more attention than the realistic-styled peer-learners (M = 55.28 sec, SD = 65.65 sec) with ($F(1,264) = 54.13$, $p < .0001$, $\eta^2 = .17$).

Furthermore, for different hand-raising manipulations, attention time on the peer-learners differs significantly with ($F(3,264) = 6.93$, $p < .001$, $\eta^2 = .07$). Particularly, the total time spent on peer-learners in the 80% condition (M = 88.95 sec, SD = 78.15 sec) is significantly longer than in the 65% condition (M = 59.23 sec, SD = 65.19 sec) with ($F(3,264) = 6.93$, $p < .0001$, $\eta^2 = .07$). In addition, the total time spent in the 20% condition (M = 88.62 sec, SD = 87.53 sec) is significantly longer than in the 65% condition (M = 59.23 sec, SD = 65.19 sec) with ($F(3,264) = 6.93$, $p = .005$). In summary, attention time towards extreme levels of hand-raising percentages are longer than for intermediate levels.

Additionally, we found some significant interaction effects regarding the attention time on the peer-learners. The time on peer-learners in the hand-raising condition depends on the sitting position of the students with ($F(3,264) = 3.88$, $p = .0097$, $\eta^2 = .041$), as well as the attention time on peer-learners in the avatar visualization styles condition depends on the sitting position with ($F(1,264) = 11.37$, $p < .001$, $\eta^2 = .039$) and vice versa. A small interaction effect was found between the hand-raising condition and the avatar visualization styles with ($F(3,264) = 3.36$, $p = .02$, $\eta^2 = .036$).

Visual Attention on Instructor

Total time spent on instructor for different sitting positions, avatar visualization styles, and various hand-raising behaviors are depicted in Figure B.4 (a), (b), and (c), respectively. The participants that are seated in the front (M = 190.07 sec, SD = 93.13 sec) attended to the virtual instructor significantly more than the participants seated in the back (M = 80.37 sec, SD = 60.78 sec) with ($F(1,264) = 144.16$ $p < .0001$, $\eta^2 = .34$).

The virtual instructor drew significantly more attention in the realistic-styled avatar condition (M = 145.98 sec, SD = 96.63 sec) than in the cartoon-styled avatar condition (M = 114.82 sec, SD = 89.83 sec) with ($F(1,264) = 11.81$, $p < .001$, $\eta^2 = .04$).

Furthermore, attention time on the instructor is found to differ significantly between different hand-raising behaviors of the peer-learners with ($F(3,264) = 3.54$, $p = .015$, $\eta^2 = .04$). In particular, the total time spent on virtual instructor in the 65% condition (M = 152.46 sec, SD = 91.48 sec) is significantly longer than the 80% condition (M = 117.39 sec, SD = 91.12 sec) with ($F(3,264) = 3.54$, $p = .009$, $\eta^2 = .04$). Overall, more attention is drawn by the virtual

instructor in the intermediate levels of hand-raising than the extreme levels. There were no interaction effects found for attention time on instructor.

Visual Attention on Screen

Total time spent on the screen, where the lecture content visualized for different sitting positions, avatar visualization styles, and various hand-raising behaviors are depicted in Figure B.5 (a), (b), and (c), respectively. The participants that are seated in the front (M = 218.65 sec, SD = 78.70 sec) attended to the lecture screen for a significantly longer period of time than the back seated participants (M = 154.21 sec, SD = 96.88 sec) with ($F(1,264) = 42.5$, $p < .0001$, $\eta^2 = .14$).

We did not find significant effects on screen attention between cartoon- and realistic-styled avatar conditions ($F(1,264) = 1.9$, $p = .17$, $\eta^2 < .01$); however, attention time in realistic style (M = 193.35 sec, SD = 92.30 sec) was slightly longer than cartoon style (M = 173.95 sec, SD = 96.11 sec).

In addition, the total attention time on the screen is found to differ significantly between different hand-raising conditions with ($F(3,264) = 5.74$, $p < .001$, $\eta^2 = .06$). In particular, attention time on screen is longer in the 65% hand-raising condition (M = 222.03 sec, SD = 94.90 sec) than in the 80% condition (M = 156.06 sec, SD = 88.25 sec) with ($F(3,264) = 5.74$, $p < .001$, $\eta^2 = .06$). In addition, attention time in the 65% condition is also significantly longer than in the 35% hand-raising condition (M = 174.87 sec, SD = 81.28 sec) with ($F(3,264) = 5.74$, $p = .025$). The overall trend of attention on the lecture screen is similar to virtual instructor with the intermediate conditions being higher than the extreme conditions. There were no interaction effects found for attention time on screen.

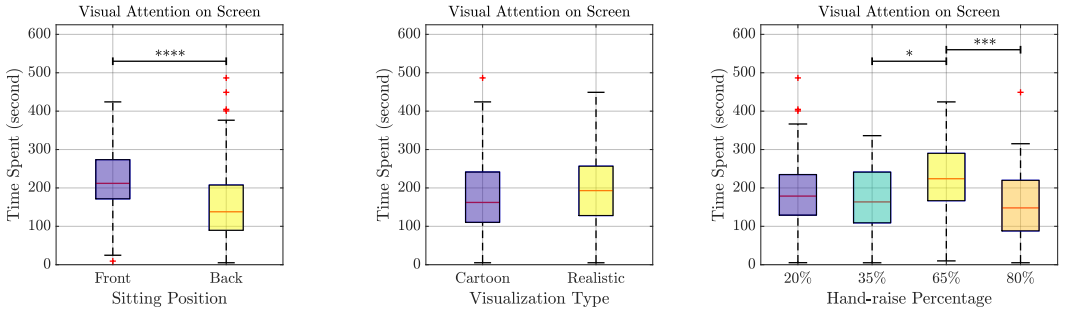
B.1.6. Discussion

We discuss experimental results particularly for social interaction and dynamics in VR classrooms, usability of eye tracking data, and the advantages of such classrooms along with their limitations.

Social Dynamics in VR Classroom

We discuss our findings about social dynamics in the VR classroom in three parts, particularly based on **H1**, **H2**, and **H3** which are related to different sitting positions, different avatar visualization styles, and different hand-raise behaviors of peer-learners, respectively.

In our analyses, we found that the participants seated in the front of the classroom attended



(a) Comparison between sitting positions.

(b) Comparison between visualization types.

(c) Comparison between hand-raising behaviors.

Figure B.5.: Attention towards screen for different classroom manipulation configurations. $*$, $***$, and $****$ correspond to the significance levels of $p < .05$, $p < .001$, and $p < .0001$, respectively.

less on the peer-learners than the participants in the back, which was expected because they had fewer peers in their FOV, unless they turn back of the classroom. Assuming that during the course of the lecture, participants are supposed to listen and pay attention to the topics told by the instructor, the visual attention we observed is normal. Briefly, this is an indication that participants focus on the lecture content or instructor instead of visually interacting with their peers when seated in the front. Further, as a supporting evidence to aforementioned result, front seated participants had spent significantly more time visually attending the instructor and the screen than the participants seated in the back. We assume that these results are due to being closer to them and having fewer occluding objects in the frontal participants' FOV. These findings confirm our **H1**. Additionally, the results from the interaction effects support this hypothesis. The differences in visual attention on their virtual peer-learners for the avatar visualization style and hand-raising depend on the sitting position. Participants located in the back of the classroom have more peer-learners in their line of sight and therefore recognize the behaviour of the virtual peer-learners more, than participants seated in the front.

Our results indicate that students visually attended for longer on the peer-learners when avatars in the classroom were presented in cartoon styles. Considering the number of peer-learners in the environment and the ages of our participants being between 10-13, we argue that participants may have felt like engaging more with their peer-learners due to the emotional reasons as cartoon-styled peers are more appropriate to their ages. Realistic-styled peer-learners may be too ordinary for student engagement with peers in our setup, which

led to less amount of attention. On the contrary, participants visually spent more time on the instructor when realistic-styled avatars were used. We conceive that if the avatar styles are ordinary, then the visual attention shifts to the instructor instead of interacting with the peer-learners. Lastly, as we did not find any statistical difference in attention time on the screen between different avatar visualization styles, we conclude that visual attention on the screen is not affected by such avatar visualization styles. Realism that is provided by the avatar styles may introduce additional computational complexity as such visualizations can be computationally expensive or can require additional effort to implement in advance. If the interaction with peer-learners is the main focus of the lecture, then practitioners can opt for cartoon-styled avatars. This also decreases the effort of generating the avatars. Overall, these findings confirm our **H2**.

In the analysis on different hand-raising behaviors of the peer-learners, we found mixed effects. In the attention time towards peer-learners, we found a clear evidence that attention time in the extreme hand-raising conditions, namely when 80% or 20% of the virtual peer-learners raise their hands after the questions were asked by the virtual instructor is longer than in the intermediate conditions (35% and 65%). The extreme conditions may represent either more or less capable groups of peer-learners in the learning environment and participants may have a higher self-concept when surrounded by a less capable group and the other way around, which is related to the Big-fish-little-pond effect [320]. Having reasonably higher attention on peer-learners on these conditions also indicates that VR can present an opportunity to create digital environments to further study students' self-concept. On the other hand, intermediate hand-raising conditions may help students to focus more on learning related objects in the classroom instead of peer-learners such as lecture content or instructor as experimentally indicated. However, we expected an approximately linear increase in terms of attention time towards higher hand-raising conditions in the attention time on peer-learners. While we obtained an expected result between the 65% and 80% hand-raising conditions, the results regarding the 20% hand-raising condition do not support our hypothesis **H3**. This might be due to a moment of surprise when only a handful of peer-learners raises their hands indicating that few number of peer-learners know the answers of the questions. Furthermore, we found that attention time on the instructor tended to be longer in the intermediate levels of hand-raising than in the extreme conditions. Statistically significant results are only found for the difference between the 65% and 80% condition. While a decreasing linear trend towards the higher hand-raising percentages exists between the 65% and 80% for attention on the instructor, the overall trend is against our hypothesis,

even though they are aligned with the attention time on peer-learners. Lastly, the experimental results on attention time on the screen is similar as compared to the attention time on the instructor. However, the 35% hand-raising condition drew significantly less attention than the 65% condition, which does not support our hypothesis. Overall, while some of our expectations are verified, **H3** is not confirmed. Still, the resulting behaviors should be further investigated with regard to effects on students' self-concepts during VR learning and considered when creating a classroom students are habituated to.

In summary, the three different manipulations that we studied have important effects on students' visual behavior in immersive VR classrooms in terms of social dynamics. For instance, in practice, students' self-concept can be affected by consistent hand-raising behaviors of virtual avatars over the time. While this may be less problematic in real classrooms as peer students may have different capabilities in different themes, it should carefully considered in the virtual setting, because we could present always the same behavior of the peer-learners. An adaptive strategy for hand-raising behaviors of the virtual peer-learners may be considered in practice. In addition, seating the students in the front along with realistic-styled avatars may help to increase visual attention on the lecture content. However, if a more interactive classroom environment is focused on visual interaction, practitioners can either seat students in locations where they can see their peer-learners clearly or design VR classrooms differently in terms of seating plans.

Usability of Eye Tracking Data

As eye tracking data is considered a noisy data source, we discuss our insights into the usability of this data, for particularly the immersive VR classroom setups. As aforementioned, we defined the visual attention on the different objects by using an attention threshold, which was 200 ms. In the end, in almost all conditions, the total amount of time that was spent on only the three types of objects was in the vicinity of half of the complete experiment duration despite having a relatively higher attention threshold value compared to fixation detection algorithms in the eye tracking literature. Such amount of total attention time on these three objects empirically validates our assumption of independence between them as well. We removed a significant number of samples from eye movement data due to sensory issues (e.g., lower eye tracking ratio) in order to obtain high-quality data and accurate attention mapping on the objects in the virtual classroom. While this may not be necessary for larger objects such as virtual screen in the classroom, it might cause mapping the attention wrongly for the smaller objects such as virtual avatars if the data quality is low. Considering that

the participants were children in our experiments and they did not have experience with virtual reality and eye tracking, number of data removals due to such issues would be more than the experiments that are carried out with adults. In addition, unlike pre- or post-tests, eye tracking allows researchers to analyze time-dependent and temporal visual behavior changes, which can help assess students' states during virtual lectures and adapt to the environment accordingly. Therefore, despite the drawbacks, we suggest using eye movement data in such classrooms as long as an accurate calibration is applied in advance. A further iteration could take relationship of eye movement-based visual attention into consideration or analyze perceived relevance of lecture content along with eye-gaze behaviors such as in [321] and [322], respectively.

Advantages and Limitations

One of the advantages of immersive VR classroom setups is the opportunity of simulating different classroom manipulations in remote settings, which are difficult to do in real world, and evaluate students' behaviors and learning under such manipulations. Another advantage of such setups is the possibility of preserving the privacy of students since the videos that include faces are not recorded in such settings. In real world classrooms, it is troublesome to record and store videos of the class while lecturing, even though there are some efforts supporting the automated anonymization [323] of such data. In contrast, data collected from virtual classrooms can be pseudo-anonymized. However, one should be aware of the amount of personal information that can be extracted from eye movement data and how to manipulate it [218], [219], [221]. Furthermore, one should take the relationship between iris texture and biometrics into account and how to preserve privacy in case eye videos are recorded and stored [324]. In addition, we observed during experiments that some of the students intended to raise their hands when seeing the hand-raising behaviors of the virtual peer-learners. While we did not record hand tracking data in our study, it is possible to accurately assess the intentions of students towards questions asked by the virtual instructor by using a hand tracker device on the HMD, which is another advantage of VR setups compared to real classrooms. Although, hand-raising is a good indicator of children's participation during a lecture, we do not know if students interpret this behaviour of their virtual peers as a sign of competence, engagement, or motivation.

Despite the advantages, there are other technical limitations regarding the use of VR classrooms. Long periods of exposure to VR lectures can lead to immense levels of cybersickness. In addition, a vast amount of HMD movement on the head may cause a drift in eye tracker

calibration, leading to incorrect sensor readings. This can affect interaction experience if gaze-aware features are included in virtual environments. These should be taken into consideration when designing a virtual classroom and lecture. Particularly, the duration of the lecture should be chosen carefully to minimize these effects.

B.1.7. Conclusion

To understand the visual attention in VR classrooms in different manipulations, we analyzed object-of-interest information based on eye-gaze. We found that participants seated in the front attended more time to the virtual instructor and the screen displaying lecture content. In addition, participants focused on the cartoon-styled peer-learners more than realistic-styled ones, whereas in the realistic-styled avatar manipulation the virtual instructor drew more visual attention. The extreme conditions of hand-raising behaviors drew more attention towards virtual peer-learners, whereas in the intermediate conditions visual attention was focused more on the instructor and screen. These findings are based on the eye movements of the participants and correspond to the social dynamics of VR classrooms such as students' self-concept or peer-learner interaction; however, such manipulations may also affect learning outcomes. While our results provide primitive but fundamental cues about how to design immersive VR classrooms by taking students' visual behaviors into account for different goals in digital teaching, effects of such manipulations on the learning outcome should be further investigated.

As future work, we plan to specifically investigate the relationship between different manipulations with temporal gaze dynamics as an immediate response to asked questions and related students' performances.

B.1.8. Acknowledgments

This research was partly supported by a grant to Richard Göllner funded by the Ministry of Science, Research and the Arts of the state of Baden-Württemberg and the University of Tübingen as part of the Promotion Program of Junior Researchers. Lisa Hasenbein and Philipp Stark are doctoral candidates and supported by the LEAD Graduate School & Research Network, which is funded by the Ministry of Science, Research and the Arts of the state of Baden-Württemberg within the framework of the sustainability funding for the projects of the Excellence Initiative II. Authors thank Stephan Soller, Sandra Hahn, and Sophie Fink from the Hochschule der Medien Stuttgart for their work and support related to the immersive

virtual reality classroom used in this study.

B.2. Using gaze transition entropy to detect classroom discourse in a virtual reality classroom

B.2.1. Abstract

This paper explores gaze entropy as a metric for detecting classroom discourse events in a virtual reality (VR) classroom. Using data from a laboratory experiment with $N = 240$ secondary school students, we distinguished between events of teacher-centered classroom discourse (question, hand raising, answer) and teacher explanation by analyzing their transition and stationary gaze entropy. Employing multi-level regression models, both entropy measures effectively discriminated between the two events and distinguished different levels of classroom participation as indicated by the degree of hand-raising by virtual students. Furthermore, using both measures in a logistic regression model, the potential of gaze entropy could be demonstrated by predicting the two events with 67% accuracy. By analyzing transition and stationary entropy, the study attempts to uncover different gaze patterns associated with learning events in a virtual classroom. The results contribute to the research and development of VR scenarios that help to simulate effective learning environments.

B.2.2. Introduction

The classroom has been understood as a central learning environment for students. Social interactions and social relationships between teachers and students, as well as between students themselves, create a dynamic of mutual learning that has been shown to contribute to students' emotional, cognitive, and academic development [90]. A virtual reality (VR) classroom can offer socially immersive learning experiences by simulating the interactive classroom discourse between animated peer learners and the virtual teacher conducting the lesson [40], [41]. Generally, classroom discourse refers to a collaborative learning process characterized by active participation and behavioral engagement among students [325], offering learning benefits for each individual learner [67], [156]. In this study, we aimed to detect students' visual attention in multiple discursive events during a lesson. We examined the extent to which participants' gaze behavior distinguished between teacher and student discursive events and evaluated the predictive value of two entropy measures. This analysis aims to provide further insights into learners' perceptions of learning-related classroom events as a driver for their learning and achievement [143]. More specifically, we utilized the concept of gaze transition entropy to investigate visual attention in terms of visual exploration

and visual attention distribution in a VR classroom.

The importance of gaze transition entropy as a metric for discerning individual gaze patterns has been well-established in prior research [141], [157], [181], [326]. Krejtz et al. [181] proposed two entropy measures (transition entropy and stationary entropy) calculated from information about the gaze duration on areas of interest (AOI) and the transitions between them. The measures reflect predictability in AOI transitions and indicate overall gaze distribution over stimuli [181]. A higher gaze transition entropy indicates more randomness and frequent gaze switching, and a higher stationary entropy indicates a more uniform distribution of visual attention over AOIs. Both measures have proven valuable in quantifying distinct gaze patterns [181], making them an apt candidate for investigating visual attention for different events within VR classrooms. In this context, VR offers an immersive experience in a standardized experimental setting that can simulate learning environments like classrooms and mimic learning-related behaviors of animated avatars [26], [69], [161], [327]. The visual behavior towards virtual avatars is especially important for younger children since they might show stronger reactions towards social cues when confronted with animated social behavior [328].

For this reason, the study utilized gaze transition entropy to investigate events of classroom discourse in a VR classroom. The study aims to investigate whether participants' gaze transition entropy indicates elements of teacher-centered discourse exhibited by virtual avatars. We focused on a subset of classroom discourse events, such as teacher questions, hand raising, and student answers. We analyzed transition and stationary entropy as two statistical measures of visual exploration and visual attention distribution. For a sample of $N = 240$ pupils, we aim to unravel gaze patterns associated with two distinct classroom events exhibited by the virtual avatars: Teacher-centered classroom discourse (teacher questions, hand raising, and student answers) and teacher explanation (teaching the lesson content). We seek to contribute to the discussion on the effective utilization of gaze entropy by investigating the explanatory power of transition and gaze entropy in detecting these two events within a VR classroom. This leads us to formulate our first research question.

R1: Can transition and stationary entropy be used to differentiate events of animated classroom discourse (teacher questions, hand raising, and student answers) and teacher explanation (teaching the lesson content) during a VR lesson?

To further explore the utility of entropy measures, we specifically focus on hand raising as a

form of student participation during classroom discourse [325]. The experiment manipulated the level of student participation indicated by the number of virtual students who raised their hands. Each participant was assigned to one of four hand-raising conditions, which allowed us to analyze the effect of different levels of hand-raising on gaze transition entropy. We can formulate the second research question by incorporating the hand-raising conditions into the analysis.

R2: Does the predictive value of the two entropy measures (transitional entropy and stationary entropy) depend on different levels of student participation indicated by hand raising?

B.2.3. Related Research

In a real classroom, students are used to focusing their attention and recognizing social and learning-related behavior from the teacher and their peer learners [196]. Such attention behavior should also be evident in a VR classroom, as children are exposed to an authentic and familiar learning environment [41]. Various previous studies have investigated VR classrooms in different contexts. While some studies concentrate on the role of the teacher in a virtual classroom [8], [11], others focus on students' attention [65], [95], social-related information [6], [9], and learning [40], [99]. Further VR classroom research considered aspects of the design of the virtual environment [38], [98] as well as the sitting position of the students [37], [66] or class size [43]. Studying visual attention through eye tracking is a prominent non-invasive technique to investigate participants' behavior during a VR experience [58]. In the context of social-related information, visual attention has been studied, for example, concerning children's reaction to social stimuli [57] or the lack of attention to faces in autism [329]. Visual attention is especially relevant in educational VR since attention influences the emotional learning processes [330]. The distribution of visual attention in VR classrooms has also been studied by exploiting object of interest information [3]. Studying visual attention using gaze entropy can further be used to investigate students' event-related behavioral changes in VR classrooms with social-related information.

The method for computing gaze transition entropy has been previously introduced by Krejtz et al. [181], [197]. Their original work modeled gaze-switching patterns as Markov chains and employed two entropy measures grounded in the theory of Shannon entropy. The utility of their approach lies in the ability to quantify two measures that can be used for statistical analysis. This approach showcased its efficacy by discerning participants' gaze

patterns during free viewing of classical art paintings [181]. Moreover, individual differences in gaze transition entropy have been pivotal in assessing task load during surgery [141] and driving [142], as well as in evaluating cognitive states such as sleep deprivation [331] and cognitive strategies during pattern recognition [138]. The method's adaptability has been demonstrated, moving beyond conventional AOI approaches, by using word spans as a different type of transition [332]. The usefulness of gaze transition entropy in real-time analysis has been demonstrated in tracking participants' cognitive states [333]. A comprehensive examination of gaze entropy within the context of visual attention is outlined by Shiferaw et al. [143]. Gaze transition entropy has also been applied in the field of education research to identify visual attention dynamics during interactive multimedia learning [157], in chemistry education [326], and to assess teacher competencies [139]. Given that visual perception is important in recognizing social interactions and interpreting social behavior [144], gaze transition entropy potentially provides a good measure to investigate students' processing of social-related information.

B.2.4. Methods

Experiment and Sample

(a) Animated virtual students.

(b) Perspective into the classroom scene.

Figure B.6.: Images of the VR classroom showing virtual students hand raising and the whole classroom.

For the analysis, we used the data from a lab experiment described in Hasenbein et al. [6], [9], which provides data from a VR classroom experiment with virtual peer learners.

In the VR laboratory experiment, participants (sixth-grade students from schools in Baden-Württemberg, Germany) entered a 15-minute simulation of a teacher-directed lesson on

computational thinking in a virtual classroom with animated peer learners. The participants were placed individually in the same VR classroom and randomly assigned centrally in the second or last row. During the lesson, a virtual teacher explained the topic and posed questions for students to engage. Virtual peer learners raised their hands in response. The teacher and the virtual students were non-playable characters with a predefined set of behaviors. The virtual teacher referred to slides on the presentation board to show the learning materials. The lesson content and the number of classroom discourse events were consistent across all configurations. A detailed timetable of the events is available in the Appendix C.

Further, the experiment employed a between-subjects design to investigate different levels of participation of the virtual students. The learning-related participation of virtual peer learners was manipulated by four levels of hand-raising. Either 20%, 35%, 65%, or 80% of the virtual peers raised their hand after a question from the teacher. Eye-tracking data was collected from all participants ($N = 381$).

The experiment used an HTC-VIVE Pro Eye head-mounted display and the integrated Tobii Eye tracker, with a trackable field of view (FOV) of 110° and a reported accuracy of 0.5°–1.1° within the 20° FOV. The Unreal Game Engine v4.23.1 was employed to render the virtual scene. An image of the animated virtual students can be seen in Figure B.6a, and a picture of the classroom taken from the back can be seen in Figure B.6b.

This specific dataset has already been investigated in previous research. The participants' visual attention has been analyzed to detect differences in classroom characteristics, such as the virtual avatars' sitting position or visualization style [3]. The visual scanning patterns of participants have also been used to analyze social comparison behavior [9] and their learning experience [6].

Data Aggregation and Measures

From the VR experiment, we obtained eye, gaze, and head information for each time frame. Using the gaze-ray casting technique, we obtained participants' gaze-intersection points with all virtual objects in the environment [5]. In the first data cleaning step, we removed all participants with a tracking ratio lower than 90% in the reported pupil diameter variable. For the remaining sample ($N = 240$), we selected a set of objects to be considered for our analysis. We defined the virtual teacher, the (presentation) board, and each animated virtual student as separate areas of interest (AOIs). This led to a total of 26 AOIs. With the selected set of AOIs, we created duration and transition datasets by calculating the duration of gaze intersection

on each AOI and the gaze transitions between them. To eliminate longer transitions, we set a maximum threshold for transition duration of 4.50 seconds (0.99-quantile). We also excluded durations on AOIs smaller than 50 milliseconds to control for the imprecision of the eye tracker (0.01-quantile). After these preprocessing steps, we calculated the AOI duration and transition matrices for 30-second intervals, with a sliding window of 10 seconds for all experiment sessions. We filled the missing entries with zero for all AOIs that did not occur during the 30 second intervals. Afterward, we normalized the transition and stationary matrices to represent maximum likelihood estimators of their theoretical probability distribution [181].

Transition and stationary entropy were calculated according to Krejtz et al. [181] for each 30-second interval. We removed the complete data of participants with more than 20% missing entropy values in at least one of the variables. We also dropped singular data points containing missing values. As a result, data from 240 participants with $N = 17202$ data points could be used for the final analysis. Given the timetable of the VR experiment, we labeled every 10 second interval as either an event related to teacher explanation or classroom discourse. Intervals with elements of classroom discourse showed a combination of teacher questions, hand raising, and student answers. Any of these combinations were coded as one, while any event containing a teacher explanation was coded as zero. The coding of the events can also be found in Appendix C. Note that the classroom discourse events only referred to the animated behavior of the virtual avatars and were not related to any participant behavior. The final binary event variable (1 =classroom discourse and 0 =teacher explanation) was then created for each of the 30-second intervals in the following way: If at least one classroom discourse event happened during the interval, it was labeled 1 (else 0). This allowed us to compare the gaze transition entropy measures calculated for 30-second intervals with the events that occurred during the same interval length.

Data Analysis

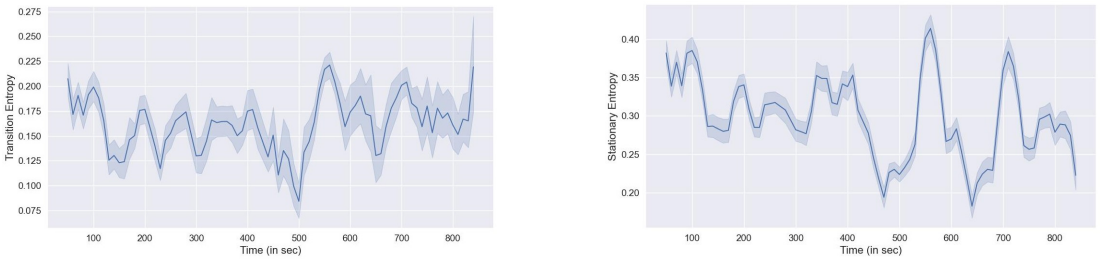
In the first analysis step, we used multi-level linear regression models that were applied separately for transition and stationary entropy. This analysis helped us explore the explanatory power of the two measures concerning classroom events and the levels of student participation indicated by hand raising. Therefore, in addition to the binary event as an independent variable, we added the hand-raising conditions as additional independent variables. We used the 35% and 65% hand-raising conditions as the reference group to which the 20% and the 80% conditions were compared. Previous research showed no difference between the reference conditions, and thus, they can be merged to represent average hand-raising [6], [9].

Furthermore, we tested for interaction effects between the event and the hand-raising variables. We modeled participants as a random intercept to prevent overestimating significance testing for the hand-raising variables.

In the second analysis step, we used a logistic regression model to predict the binary event variable (classroom discourse or teacher explanation). Only the two entropy measures were used as independent variables in the model. We applied a person-mean centering on both entropy measures by subtracting participants' mean entropy from each entropy value of their experiment session. The dataset was randomly split by a 80 : 20 ratio. For 50 iterations, we reported mean accuracy, f1 score, and an average confusion matrix for the model predictions on the test sets.

B.2.5. Results

For all time intervals ($N = 17202$), transition and stationary entropy were correlated with $r = 0.17$. Descriptive statistics revealed differences in mean values between the events for transition entropy (classroom discourse: $M = 0.18$, $SD = 0.13$, teacher explanation: $M = 0.15$, $SD = 0.18$) and stationary entropy (classroom discourse: $M = 0.34$, $SD = 0.13$, teacher explanation: $M = 0.26$, $SD = 0.11$). The changes over time for transition entropy are displayed in Figure B.7a and for stationary entropy in Figure B.7b.



(a) Average transition entropy.

(b) Average stationary entropy.

Figure B.7.: Time curve of average entropy measures (mean and standard deviation) of all participants during the full experiment.

In the first step of the analysis, two multi-level linear regression models were analyzed with either transition entropy or stationary entropy as the independent variable. Results from the transition entropy model (see Table B.1) showed a significant positive increase for the classroom discourse event (Estimate: $\beta_{Event} = 0.20$, $p < 0.001$). The predictive value of transition entropy was higher in the 20% and in the 80% hand-raising condition compared to

the average hand-raising conditions ($\beta_{20\%} = 0.14$, $p = 0.03$ and $\beta_{80\%} = 0.17$, $p = 0.01$). There was no significant interaction effect between the event and the hand-raising condition.

Results from the stationary entropy model (see Table B.2) showed a significant positive increase for the classroom discourse event ($\beta_{Event} = 0.63$, $p < 0.001$). The predictive value of stationary entropy was higher in the 80% condition than in the average hand-raising conditions ($\beta_{80\%} = 0.21$, $p = 0.01$). There was no significant change for the 20% hand-raising condition or any interactions. Figures for the time curve of both entropy measures separated by the hand-raising condition are shown in Appendix B.

Table B.1.: Results of the multi-level linear regression analysis with transition entropy as the dependent variable. Event represents the binary event variable. The hand-raising variables indicate participants' assignment to the respective experimental condition.

Model summary			N obs = 17202	
Dep. Var.: Transition entropy			N groups = 240	
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.19	0.04	-4.78	0.00
Event	0.20	0.02	9.34	0.00
Hand-raising 20%	0.14	0.06	2.15	0.03
Hand-raising 80%	0.17	0.06	2.69	0.01
Event × Hand-raising 20%	-0.01	0.03	-0.34	0.74
Event × Hand-raising 80%	0.02	0.03	0.61	0.54

Table B.2.: Results of the multi-level linear regression analysis with stationary entropy as the dependent variable. Event represents the binary event variable. The hand-raising variables indicate participants' assignment to the respective experimental condition.

Model summary			N obs = 17202	
Dep. Var.: Stationary entropy			N groups = 240	
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.44	0.05	-8.31	0.00
Event	0.63	0.02	34.21	0.00
Hand-raising 20%	0.13	0.08	1.57	0.12
Hand-raising 80%	0.21	0.08	2.47	0.01
Event × Hand-raising 20%	0.01	0.03	0.24	0.81
Event × Hand-raising 80%	0.03	0.03	1.06	0.29

In the second step, we used a logistic regression model to analyze the predictive power of both entropy measures for differentiating classroom discourse events. The logistic regression model, using transition and stationary entropy as input variables, predicted the two events correctly with 67% accuracy ($f1 = 0.67$). The number of false-positive and false-negative samples was balanced for both events. All model details can be found in Table B.3. Since previous research used a smaller number of total AOIs [332], we created an additional dataset merging all student-related AOIs into one AOI. We also repeated the entire data processing and analysis for comparison purposes with only 3 AOIs (student, teacher, and board). The results showed smaller predictive power when using only 3 AOIs with an accuracy of 61%. Details on the analysis can be found in Appendix A.

Table B.3.: Results of the logistic regression model predicting events of classroom discourse (Class. Discourse) and teacher explanation (Teach. Expl.). Mean-centered entropy measures were used to predict the classes. Accuracy, f1 score, and the confusion matrix are reported as mean values or in percent over 50 random-split iterations (test size 0.2).

Model performance		$N = 17202$
Mean accuracy: 0.67		Class. Discourse: $N = 9009$
Mean f1: 0.67		Teach. Expl.: $N = 8193$
Confusion matrix (in percent)		
	Predicted Class. Discourse	Predicted Teach. Expl.
True Class. Discourse	0.37	0.16
True Teach. Expl.	0.17	0.30

B.2.6. Discussion

In our study, we investigated whether gaze transition entropy can be used to detect events with elements of classroom discourse in a VR classroom, utilizing the large dataset comprising 240 participants. Transition and stationary entropy revealed differences for events containing elements of teacher-centered classroom discourse (teacher question, hand raising, student answer) compared to events of teacher explanation (R1). When manipulating the number of hand-raising exhibited by the virtual students, the results revealed a higher predictive value of transition entropy for the 80% and 20% hand-raising conditions compared to average hand-raising. This indicated that participants showed more visual exploration in both of these conditions compared to conditions of average hand raising (35% and 65%). Because both

of these conditions provided relevant information for learning (everyone or no one wants to participate or knows the answer), they may have triggered stronger exploratory behavior. Stationary entropy only revealed a higher predictive value in the 80% hand-raising condition. This indicated that participants' attention was more uniformly distributed only when many virtual students engaged in the classroom discourse (R2). However, no interaction was found between the event variable and the conditions, suggesting that this gaze-related behavior was not only present during the events.

The successful application of gaze transition entropy has proven effective in detecting events of classroom discourse, relying on observable indicators like teacher questions, hand raising, and student answers. However, it's important to note that these indicators only represented a subset of the diverse interactions occurring during classroom discourse. The experiment intentionally focused on this specific subset, specifically targeting learning-related elements of a teacher-centered discourse. The chosen indicators were deliberately selected for their clarity and ease of observation. The present research was only a first step in identifying and understanding student reactions during classroom discourse. It's crucial to acknowledge the study didn't aim to explain the entirety of complex interactions in a classroom. Instead, we focused on clear and distinctive events. Future experiments could overcome this limitation by incorporating more detailed discursive events within VR classrooms, offering a more comprehensive understanding of the nuanced dynamics at play.

One notable strength of this study was its highly standardized setting, where participants observed identical lessons featuring the same animations, avatars, and elements of classroom discourse. This standardized environment ensured consistency across the experimental conditions, facilitating precise analysis and comparison of gaze transition entropy. However, this also represented a limitation with regard to the complexity of the classroom situation and the participants' possibility to interact. To ascertain the reliability of gaze transition entropy, future investigations should extend beyond the controlled virtual setting to examine the complexity of real classroom scenarios. Mobile eye trackers could prove instrumental in capturing dynamic social interactions during a real lesson, providing valuable insights into the applicability and robustness of the measures.

We also encountered technical challenges investigating gaze transition entropy in a 3D virtual environment. In contrast to previous research, we did not use a grid-based approach to calculate the entropy measures. The characteristics of 3D virtual environments allowed us to obtain AOI duration and transitions directly based on the gaze intersections with the virtual objects. The advantage of this approach was that no fixation detection algorithms

needed to be applied. However, this approach also imposed some limitations. First, given the accuracy and precision of the VR eye tracker, participants' gaze directions may have been impaired. This could have led to an underestimation of AOI duration and transitions. Second, selecting AOIs in VR based on virtual objects resulted in undetected gaze transitions between other objects in the environment. While the selection of specific AOIs has the potential to improve the specificity of the visual attention analysis, it also led to a loss of information and a higher number of missing or zero transition entropy values. Additionally, the prediction accuracy diminished when only 3 AOIs (student, teacher, and board) were used to calculate the measures. This indicated that 3 AOIs provided insufficient information when all student-related transitions were merged. Another challenge was selecting a suitable time interval. Although one event occurred every 10 seconds, the amount of missing data for 10-second intervals forced us to use longer time intervals. For intervals smaller than 30 seconds, often no gaze intersection or no transition occurred for the selected AOIs, resulting in missing entropy values.

Despite these limitations, the accuracy of the predictive model using only two measures underscored the significant explanatory power of gaze transition entropy. The measures successfully predicted the events within the VR classroom scenario and further revealed differences in the participation levels of the virtual students. Future research could enhance the model's predictive power by integrating additional measures. Furthermore, our results suggested that students exhibited distinct behavior with respect to animated social avatars in VR. This highlights the importance of design aspects of virtual avatars in educational environments and emphasizes the need for further research to optimize the effectiveness of these elements in facilitating VR learning experiences.

B.2.7. Conclusion

The presented study leveraged gaze transition entropy as a valuable metric for detecting elements of classroom discourse in virtual reality (VR) classrooms by analyzing eye-tracking data of 240 participants. The analysis revealed differences in transition and stationary entropy for events related to teacher-centered classroom discourse, specifically teacher questions, hand raising, and student answers. Both transition and stationary entropy measures were found to be instrumental in distinguishing gaze patterns during discursive events compared to events of teacher explanation. These findings are a first step in exploring visual attention during virtual classroom discourse and emphasizing the impact of social avatars when designing effective VR learning environments.

B.2.8. Acknowledgements

This research was supported by a grant to Richard Göllner funded by the Ministry of Science, Research and the Arts of the state of Baden Württemberg and the University of Tübingen as part of the Promotion Program for Junior Researchers. Philipp Stark is a doctoral student at the LEAD Graduate School & Research Network, which is funded by the Ministry of Science, Research and the Arts of the state of Baden-Württemberg within the framework of the sustainability funding for projects from Excellence Initiative II. We want to thank Jens-Uwe Hahn, Stephan Soller, Sandra Hahn, and Sophie Fink from the Institute for Games, Department of Computer Science and Media at the Hochschule der Medien Stuttgart for their extensive work preparing the immersive virtual reality classroom used in this study.

C. Gaze-based Networks and Learning with Simulated Classmates

The following publications are enclosed in this chapter:

[5] **P. Stark**, L. Hasenbein, E. Kasneci, and R. Göllner, "Gaze-based attention network analysis in a virtual reality classroom", *MethodsX*, vol. 12, p. 102 662, Jun. 1, 2024. DOI: 10.1016/j.mex.2024.102662

[6] L. Hasenbein, **P. Stark**, U. Trautwein, A. C. M. Queiroz, J. Bailenson, J.-U. Hahn, and R. Göllner, "Learning with simulated virtual classmates: Effects of social-related configurations on students' visual attention and learning experiences in an immersive virtual reality classroom", *Computers in Human Behavior*, vol. 133, p. 107 282, Aug. 1, 2022. DOI: 10.1016/j.chb.2022.107282

C.1. Gaze-based attention network analysis in a virtual reality classroom

C.1.1. Abstract

This article provides a step-by-step guideline for measuring and analyzing visual attention in 3D virtual reality (VR) environments based on eye-tracking data. We propose a solution to the challenges of obtaining relevant eye-tracking information in a dynamic 3D virtual environment and calculating interpretable indicators of learning and social behavior. With a method called "gaze-ray casting," we simulated 3D-gaze movements to obtain information about the gazed objects. We used this information to create graphical models of visual attention, establishing attention networks. These networks represented participants' gaze transitions between different entities in the VR environment over time. Measures of centrality, distribution, and interconnectedness of the networks were calculated to describe the network structure. The measures, derived from graph theory, allowed for statistical inference testing and the interpretation of participants' visual attention in 3D VR environments. Our method provides useful insights when analyzing students' learning in a VR classroom, as reported in a corresponding evaluation article with $N = 274$ participants.

- Guidelines on implementing gaze-ray casting in VR using the Unreal Engine and the HTC VIVE Pro Eye.
- Creating gaze-based attention networks and analyzing their network structure.
- Implementation tutorials and the Open Source software code are provided via OSF: `https://osf.io/pxjrc/?view_only=1b6da45eb93e4f9eb7a138697b941198`.

C.1.2. Method Details

Background and Motivation for Applying the Method

Due to recent technological innovations, virtual reality (VR) has celebrated a rebirth in the consumer market, with immersive, head-mounted VR devices at affordable prices applicable in different fields of all our lives [10]. Specifically, recent developments in hardware, software, and design have resulted in VR applications being more frequently used in education and education research [19]. With VR, learning environments like virtual classrooms can be studied systematically, for example, to investigate classroom complexity [334], seating

Subject area	Computer Science
More specific subject area	Human-Computer Interaction, Virtual Reality, Eye Tracking
Name of your method	Gaze-based Attention Network Analysis
Name and reference of the original method	Ray Casting, Network Analysis
Resource availability	**Hardware:** (1) HTC VIVE Pro Eye `https://www.vive.com/us/product/vive-pro-eye/overview/` **Software:** (1) Unreal Engine `https://www.unrealengine.com/de` (2) SRanipal Unreal SDK `https://developer.vive.com/resources/vive-sense/eye-and-facial-tracking-sdk/documentation/` (3) Python 3.11 + any IDE (4) Python package requirements (see requirements.txt): • numpy https://numpy.org/ • pandas https://pandas.pydata.org/ • networkx https://networkx.org/ (5) Eye-tracking C++ scripts (see OSF) (6) Analysis pipeline in Python (see OSF) `https://osf.io/pxjrc/?view_only=1b6da45eb93e4f9eb7a138697b941198` For an illustration of the VR environment and original experiment, see `http://vre-tuebingen.de`.

Table C.1.: Specification Table

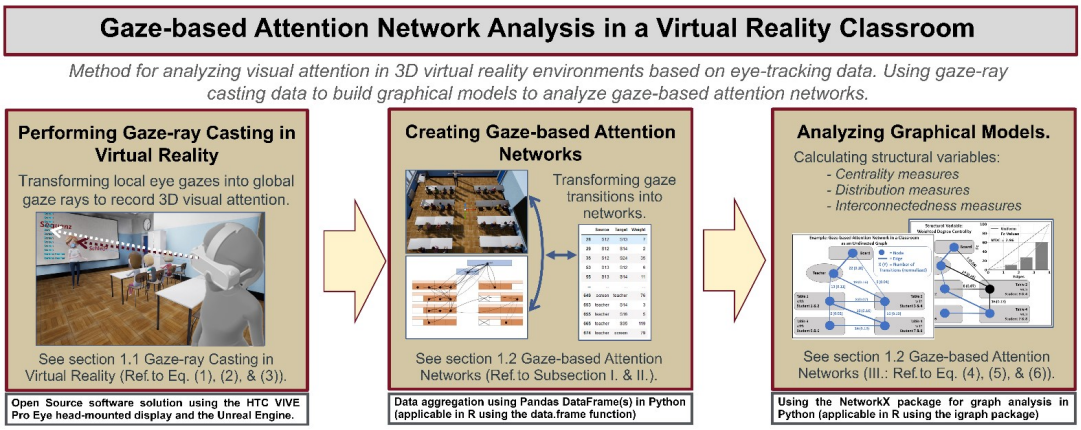


Figure C.1.: Graphical abstract

arrangements [66], or performance-related classroom behavior [6]. These developments embrace the possibility of uncovering aspects of learning that were previously difficult to study in these settings, such as visual attention. However, collecting and analyzing visual attention information in VR poses some challenges. A significant challenge involves acquiring relevant, high-quality eye-tracking data in dynamic 3D virtual environments. Aggregating this data to obtain meaningful indicators of learning and social behavior presents another complex challenge.

This method article presents a way to overcome these challenges and provides systematic step-by-step guidelines for measuring and analyzing human visual attention based on VR eye-tracking data. Our article is published alongside a corresponding empirical evaluation by Hasenbein et al.[6], where we evaluated our method with N=274 students to investigate learning with simulated virtual classmates in a VR classroom. We examined gaze-based visual attention to investigate students' performance, interest, and self-concept during a 15-minute teaching unit in a virtual classroom. The visual attention of these participants was analyzed based on eye movement data. More specifically, eye movements were obtained and analyzed using a VR device with a head-mounted display (HMD) and an integrated eye tracker [330]. For our experiment, the HTC VIVE PRO EYE head-mounted display (HMD), which operates with the Tobii eye tracker [335], was used. The Unreal Engine allows game developers to provide an immersive, interactive, and animated visualization of virtual learning environments for application in education research [19].

Given our VR experiment, the first challenge occurred when collecting eye-tracking infor-

mation. In our study, we specifically focus on overt visual attention, the process of attentional shift using eye movements [114], which can be measured by simulated gaze movements in space. More specifically, overt visual attention towards an object is the intersection of a human's gaze ray with an object for a specific time. The integrated HMD eye tracker only reports participants' local gaze directions, which does not consider the position and viewing direction. To analyze the focus points of visual attention and collect semantic information about the gazed objects, HMD, and eye tracker data must be combined and processed. This can be achieved by applying gaze-ray casting, where humans' gaze direction is transported in the environment to observe where it hits. This method already exists for other game engines and eye-trackers. A more detailed review of this challenge can be found in Ugwitz et al. [58], [73], [75]. We faced the challenge that a simple software approach was missing for the Tobii Eye Tracker in combination with the Unreal Engine. Therefore, in the first part of our article, a software solution is provided that can be easily integrated into existing projects.

The second challenge occurred when deciding on the appropriate level of data aggregation. Experimental psychology offers a wide range of methods for processing eye-tracking data when investigating learning. This is usually based on determining the smallest eye movements, such as saccades and fixations, which provide insight into underlying cognitive processes [131]. However, this method does not include information about the environment or participants' guidance of visual attention. To incorporate semantic scene information into the analysis [105], methods like scan path analysis can be applied [114], [198], [199]. However, scan path analysis relies on specific distance measures or machine learning algorithms to compare their structures [114]. Another promising way to analyze this information can be achieved by creating graphical models [200], [336], [337], which we call gaze-based attention networks. The idea of using network representations is probably most prominent in social network analysis [338], [339]. Our gaze-based networks represent participants' gaze transitions between different virtual entities in the environment over a period of time. The structure of these networks follows the mathematical principles of graph theory [182], [340], with objects in the environment treated as network nodes and gaze transitions treated as edges between them. This method has been applied for stationary eye tracking on a screen in previous research [136] concerning experimental and clinical psychology [182], [338], [340], mathematical problem-solving [136], or joint attention [148]. It allows to describe the composition and interconnectedness of the gaze-based networks using measures from graph and network theory. These measures, which we refer to as structural variables [183], [341], allow us to describe the network structure of participants' visual attention in 3D VR

environments and statistically analyze and compare participants' visual attention.

To provide a rationale for the method, the following potential advantages of the approach can be highlighted. An Open Source software solution was coded for the Unreal Engine to quickly integrate it into existing projects by following the guidelines. Also, details are presented on how the data collection pipeline with gaze-ray casting can be extended and adjusted to the needs of specific projects. Our method of transforming gaze-based information into a transition network eliminates the need to compute eye movement events such as fixation and saccades, which can be challenging in 3D environments [16], [342]. Besides minor data exclusion steps, the collected eye-tracking information can be processed without extensive data cleaning. This provides a quick way to aggregate gaze networks directly from the data collected by the gaze-ray casting pipeline.

Further, modeling eye-tracking data as gaze-based attention networks could be more intuitive to interpret for applied researchers since they can be easily visualized. Data aggregation comes with information loss and determines the possibilities of analyzing and interpreting the data, so the level of data aggregation must be chosen appropriately for the research interest. The network structures contain semantic information, including participants' reactions to virtual social actors. Especially on this level of visual attention aggregation, gaze transitions between social actors can provide meaningful information. Since empirical studies in the social sciences are interested in interpretable measures, structural variables offer a valuable and comprehensible way for statistical testing.

Structure of the article

This article is structured in four parts: First, guidelines and instructions are provided on implementing gaze-ray casting using the Unreal Engine to record gaze target information from users during a virtual reality (VR) experience (see Section C.1.2). Second, we show how to transform the obtained gaze target information into a gaze-based attention network, compute structural variables of the networks, and interpret them in the case of visual attention (see Section C.1.2). Third, the performance of the data pipeline in Python was evaluated, and samples of the code in the programming language R were provided. We hope that this can increase the applicability and reproducibility of the method, especially for researchers not familiar with Python (see Section C.1.2). Last, some general considerations for implementation and application were provided (see Section C.1.2). Additional lessons learned during the implementation are described in the *tips for application*.

Further instructions, illustrations, and implementation details are given at OSF: `https://osf.io/pxjrc/?view_only=1b6da45eb93e4f9eb7a138697b941198`. The OSF

repository structure corresponds to the article structure, with additional information within the Readme.md in each section folder. To facilitate the reproducibility of the method, the code locations are referred to throughout the text by referencing the OSF project (Ref. to OSF). Special technical terms used in the method are explained in Table C.2.

Time point	Game engines work with a specific framerate in which they update the environment. A time point is one tick or update frame in the virtual environment, considering that time intervals between two points do not differ significantly. The tick rate is based on device performance (on average, every 20ms).
Local gaze direction	A normalized vector of the HMD eye-tracker is expressed in the coordinates of the local coordinate system of the VR headset.
Global gaze direction	A vector that starts at the cyclopean eye and points into the virtual environment. This vector is stated in unreal units (uu), equal to 1 centimeter in real-life distance.
Gaze target	The virtual object is hit by the lengthened global gaze direction where the gaze position is currently located (stated in uu).
Object of interest (OOI)	Closely related to the term Area of Interest (AOI), which describes a segment of a stimulus space. OOIs are the objects of a preselected set of potential gaze targets considered in the analysis.
Gaze-ray casting	A technique to obtain gaze target (information) using the global gaze direction and object location provided by the Game Engine. See Section C.1.2 Gaze-ray Casting in Virtual Reality for detailed information.
Gaze transition	A gaze shift between two successive OOIs. More precisely, the gaze movement between the last detected gaze location on one object and the first detected gaze location on the next.
Player / User	When describing functions and algorithms in the Unreal Engine, the player is used to describe the virtual character created in the 3D space as a projection of the user's position in the room. User refers to the person who is using the VR device.

Table C.2.: Overview and explanation of technical terms used in this article.

Gaze-ray Casting in Virtual Reality

Ray casting is known and used primarily as an interactive technique in VR environments for target selection with a controller [195]. Gaze-ray casting is based on a similar idea: the direction of a human's gaze is considered a ray. Starting at the position of the cyclopean eye, the middle point between both eyes, the gaze is projected into the virtual environment,

where it hits a specific location or, in other words, a gaze target [179], [343]. The gaze-ray casting technique detects the gaze target directly during the VR experience. It enabled us to collect various information, like the label or position of the gaze target or the distance between the player and the target [3].

This has an advantage compared to remote or real-world mobile eye trackers. In remote eye trackers, gaze targets must be annotated separately by labeling the pictures or videos on the screen. This is even more complicated in real-world mobile eye trackers because of the user's free movement in a 3D space (see [344]). In contrast, when using an immersive VR [14], a 3D environment experienced through a head-mounted display, the game engine renders all of the virtual scenery. This means complete information about objects' location and shape is always available. Therefore, the gaze target is just an intersection of humans' gaze rays with the polygon surface of the closest object in the virtual space, referred to as the gaze intersection point [345]. The gaze-ray casting technique is also independent of detecting and calculating eye movement events, like fixations and saccades [342]. The only information received is about which object a user is looking at at a time point, and no eye movements need to be calculated.

The implementation of gaze-ray casting in VR with the Unreal Engine (UE) can be divided into five steps, which should be performed sequentially. In addition to an implementation tutorial (Ref. to OSF: `1-1_Gaze-rayCastinginVirtualReality/Readme.md`), a detailed description of each step is given below.

I. Enabling eye tracking in the Unreal Engine using the SRanipal SDK.

II. Creating an "eye-tracking" Actor to collect the local gaze vectors.

III. Transforming local gaze directions into global gaze directions.

IV. Projecting the global gaze vector into the environment using a ray casting function.

V. Collecting gaze target information in the eye-tracking Actor and saving it in a data file.

I. Enabling eye tracking in the Unreal Engine using the SRanipal SDK

To collect eye and gaze data with the HTC Vive Pro Eye in the UE, the provided VIVE software was used [346]. As described in the website's documentation, the SDK was integrated into an Unreal project and enabled access to all eye-tracking variables recorded by the integrated eye tracker.

II. Creating an 'eye-tracking' Actor to collect the local gaze vectors

A combination of C++ scripts and Unreal Blueprints was used to create a data collection pipeline for the project. Unreal blueprints are node-based interfaces to create gameplay elements in UE that grant easy access to already implemented functions. To further process the gaze data, a new C++ Actor Class (Ref. to OSF: `1-1_Gaze-rayCastinginVirtualReality/EyeTracker.h/.cpp`) and a corresponding Actor Component Blueprint ('BP EyeTracker') were created, where all necessary calculations were implemented. In the C++ files, the local gaze directions were stored and transformed into UE vector objects to be further processed as gaze vectors in our EyeTracker Blueprint class (Ref. to OSF: `1-1_Gaze-rayCastinginVirtualReality/Readme.md`).

Tips for application: Whenever an Unreal-type vector is created in a C++ script, the variable can be accessed in the blueprint of the connected Actor. The eye-tracking data collection from the eye tracker could also be integrated into already existing Actors in the virtual environment, but the separate data collection Actor was a convenient way to add eye tracking into already existing projects.

III. Transforming local gaze directions into global gaze directions

Continuing in the blueprint component, the location and orientation of the EyeTracker Actor had to be aligned with the player's head location and orientation. Therefore, the EyeTracker Actor was aligned with the Pawn, the main Actor in UE.

As a next step, the local gaze vector, received from the C++ script, was transformed into a global gaze direction. The forward head direction of the player could be accessed by recording the player's perpendicular head vector. This normal vector is pointing forward perpendicular to a plane describing the front or face of the player. When a user moves their head, this vector moves in sync. Simultaneously, this forward vector represented the head direction of the HMD headset but also the x-axis of the local (eye tracker) coordinate system.

Tips for application: In the blueprint, the function <<Get Forward Vector>> was used to get the forward vector with the player rotation as input.

The forward vector had to be rotated to align locally with the local gaze vector. To perform the rotation, yaw (i.e., the head rotation angle in degree to the left or right from a vertical axis) and pitch (i.e., the angle in degree at which one is looking up or down) were calculated. This method could be used because angle-based rotations are independent of the coordinate system and its units (see Figure C.2).

Euclidian geometry was used to calculate the yaw and pitch angle. As a reference vector,

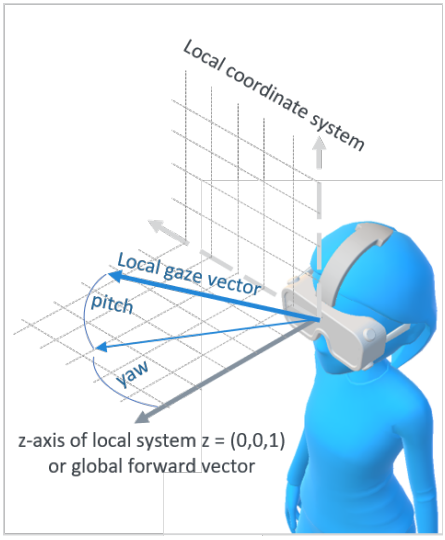


Figure C.2.: How to calculate pitch and yaw using the local gaze vector from the local coordinate system of the Tobii Eye Tracker.

the forward vector $f = (x_f, y_f, z_f) = (0,0,1)$ was used given in unreal units (uu), where 1uu is equal to 1 centimeter. With the normalized gaze vector $g = (x_g, y_g, z_g)$ and the flat 2D gaze vector $g_{flat} = (x_g, 0, z_g)$, the yaw angle in degree was calculated by

$$yaw = -cos^{-1}\left(\frac{z_g}{\sqrt{x_g^2 + z_g^2}}\right) \cdot \frac{180}{\pi \cdot sgn(x_g)} \qquad \text{with yaw} \in [-180°, 180°]. \tag{C.1}$$

The minus one and the signum function in Eq. C.1 introduced a change in the orientation of the coordinate system. The pitch angle was calculated by

$$pitch = cos^{-1}\left(\frac{x_g^2 + z_g^2}{\sqrt{x_g^2 + z_g^2}}\right) \cdot \frac{180}{\pi \cdot sgn(y_g)} \qquad \text{with pitch} \in [-180°, 180°]. \tag{C.2}$$

With both angles from the Eq. C.1 and C.2, a vector rotation (<<RotateVector>>) was performed on the forward vector f to create $f_{rotated}$. Lastly, the global gaze vector was

computed starting at the players' head location $h_{location}$ in uu as

$$g_{global} = h_{location} + (f_{rotated} \cdot k), \qquad \text{with } g_{global} \in \mathbb{R}^3 \text{ and } k \in \mathbb{R} \tag{C.3}$$

In Eq. C.3, k represents the length of the gaze vector in uu. The global gaze vector g_{global} is then also stated in unreal units. After calculating the global gaze direction, this vector was used to perform gaze-ray casting.

Tips for application: Independently of the environment, the value of k in Eq. C.3 can be set very large (we set k = 25000 uu for our experiments) because the ray cast will stop when it hits the first object. One problem when working with the Tobii eye tracker and the UE was that these two software presented gaze and head direction on different spatial coordinate systems. The Unreal Engine has a left-handed coordinate system, with positive x pointing forward (clockwise roll rotation), y pointing right (clockwise pitch rotation), and z pointing upwards (counterclockwise yaw rotation) [347]. In contrast, gaze information of the Tobii eye tracker was given in a right-handed coordinate system with z pointing forward, x pointing to the left, and y pointing up. The presented formulas for calculating yaw and pitch (Ref. to Eq. C.2 and C.3) already include the coordinate change, so pitch and yaw could be used directly for vector rotation.

IV. Projecting the global gaze vector into the environment using a ray-casting function

To get information about gaze location and gaze target, functions from the Kismet System Library were used. To perform the gaze-ray casting, either <<LineTraceByChannel>> or <<LineTracForObjects>> can be used. The two functions differ only in considering different object types as hit objects. Both blueprint functions perform ray casting automatically by taking a starting position, namely the player's head location, and an end position, namely the global gaze vector (g_{global}). Useful output variables of these functions are the name of the gaze target (Hit Component, i.e., a virtual object as a string), the 3D location of the gaze hit (Location or Impact Point in uu), and the distance from the player to the hit object (Distance in uu).

Tips for application: One important aspect is that any line trace function only returns the first hit object. Therefore, one needs to ensure that no other (potentially invisible or hidden) objects are in the player's line of sight. To this end, all colliders of hidden objects had to be disabled, while at the same time, collision for all objects one wanted to track had to be enabled such that the line trace could hit our Objects of Interest (OOIs).

V. Collecting gaze target information in the eye-tracking Actor and saving it in a data file

The gaze-ray casting output variables were stored in the blueprint for each time point and accessed via the C++ script. Together with other eye-tracking information from the integrated eye tracker, the gaze-ray casting variables (gaze location, gaze target, ray distance) were saved into a data frame marked with the timestamps. The resulting dataset was stored as a CSV file at a predefined project location.

Tips for application: To follow our code structure: In the C++ file (Ref. to OSF: `1-1_Gaze-rayCastinginVirtualReality/EyeTracker.cpp`), the gaze vector was created at line 71. Then, the gaze-ray casting was performed in the blueprint, and its output was stored starting from line 103.

Gaze-based Attention Networks

The previous pipeline collected gaze target information from users during the VR experiences. The obtained information (gaze target and time stamp) could then be used to analyze data via very different means. In our virtual classroom study [6], the gaze target information was transformed into networks, providing a flexible approach to analyze and visualize gaze-based visual attention [337], [348]–[350]. The gaze-based networks contained aggregated information about participants' visual attention in the virtual environment, represented in a network structure. Thus, we call them gaze-based attention networks. Concepts from mathematical graph theory and network analysis were used to calculate descriptive variables that reveal information about the network structure [340], which we call structural variables. The networks represented by the structural variables were then associated with social comparison and learning by performing statistical inference testing [185].

In the virtual classroom study, participants spent 15 minutes listening to a lecture about computational thinking. Gaze targets were objects in the environment from which the gaze-ray casting information was collected. One visual attention network was built per participant. Each network structure consisted of nodes, which were the virtual peer learners, the teacher, and the board in the classroom (our OOIs). The nodes were connected by edges, representing the frequencies of participants' gaze transitions between the OOIs. The more often a participant switched visual attention from one OOI to another, the larger the edge weight between two OOIs, indicating a stronger connection. As a result, each network represented a bidirectional, weighted graph of overt visual attention distribution in a virtual classroom. Example networks for two participants are shown in Figure C.3. Performing network analysis with VR gaze data and computing structural variables required three (pre-)

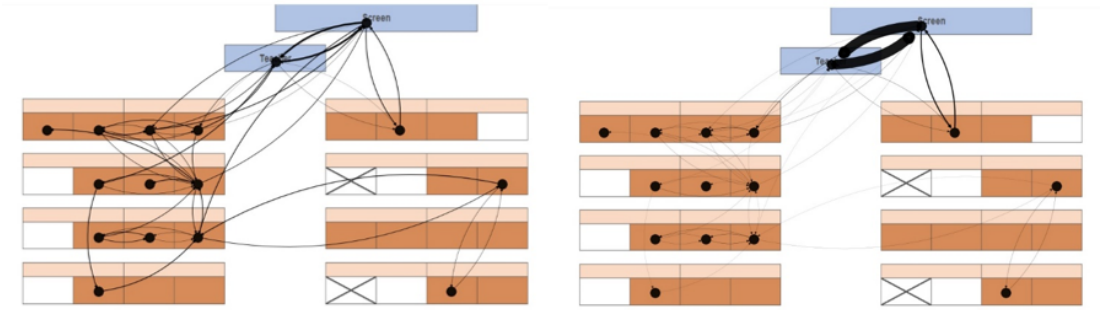


Figure C.3.: Visual representation of gaze-based attention networks from two participants in a top-down view on the virtual classroom. All OOIs are the teacher and board in blue and the positions of the virtual peer learners at their table in orange. Frequencies of gaze transitions between gazed OOIs are illustrated by the line width of the edges.

processing steps described in the following sections:

I. Aggregating raw gaze-target information into gaze transition datasets.

II. Creating gaze-based attention networks from gaze transition datasets.

III. Computing structural variables to describe gaze-based attention networks.

I. Aggregating raw gaze-target information into gaze transition datasets

As a first step, the data had to be cleaned. This can be done using any reliable procedure and considering other eye-tracking variables like pupil size to identify missing data and artifacts (e.g. [281], [282], [344]) (Ref. to OSF: `1-2_GazeBasedAttentionNetworks/fl1_preprocessing.py`). To conduct the network analysis, as explained in the subsequent steps, only the time and gaze target variables were needed.

Tips for application: The gaze-ray casting pipeline always reports a gaze target as long as a valid head direction is recorded. It is important to exclude missing values coded as placeholders for missing values (like -1). The calculation of pitch and yaw is also performed with incorrect values. If outliers and missing data are not excluded in a separate step, the global gaze direction might be incorrect, and the gaze target might be false.

After cleaning the raw data, a new data frame was created for each participant, which consisted of all gaze transitions between OOIs during one experimental session. Changes in OOIs were stored in the new transition dataset by iterating through each row of the raw

participant	time	time_dur	Source	Target
58	0.619	0.333	S14	S35
58	1.352	1.066	S35	S16
58	2.707	0.023	S16	screen
58	3.372	0.023	screen	teacher
58	4.019	0.021	teacher	screen
...	...	...	...	...
58	842.078	0.022	screen	S35
58	842.145	0.112	S35	S14
58	843.235	0.021	S14	S35
58	849.563	0.022	S35	screen
58	849.652	0.267	screen	S34

Figure C.4.: Transition data frame, with transitions between starting and landing OOI (form Source to Target), marked with the starting time of the transition and the transition duration. The participant variable indicates that such a data frame is created separately for each participant.

dataset. Thus, if the gaze target was not the same as in the previous line, the following information was stored in the transition dataset: the time stamp at the beginning of the transition, the transition duration, the transition starting object, and the transition landing object. An example of a transition data frame is shown in Figure C.4.

As a last step, some transitions needed to be excluded. We decided on an upper threshold for the maximum transition duration. The time duration variable was used to filter the dataset for longer durations (Ref. to OSF: `1-2_GazeBasedAttentionNetworks/fl2_transition.py`).

Tips for application: Excluding longer transitions seems necessary to ensure that only direct transitions are counted. If a transition duration was too long, it was likely that no direct shift was observed but rather many shifts between objects that were not considered OOIs.

II. Creating gaze-based attention networks from gaze transition datasets

Now, the datasets were used to aggregate the input variables necessary to create graph objects with the networkx [351] package. This Python package offers useful default functions, which can be customized for later analysis. With <<networkx.from_pandas_edgelist()>>, a graph object was created directly from an adjacency-like data frame. This required creating a new dataset containing a source and target variable. In this new dataset, the source and target variables stored information about the connected nodes, while a third (weight) variable contained information about the strength of the edge connection. The weights were calculated to describe the total number of gaze transitions between respective OOIs. The weight variable held all edge information from a weighted graph by counting the total

Source	Target	Weight
S12	S13	7
S12	S14	2
S12	S24	35
S13	S12	6
S13	S14	11
...	...	...
screen	teacher	76
teacher	S14	3
teacher	S16	5
teacher	S35	119
teacher	screen	79

Figure C.5.: An adjacency-like pandas edge list data frame. Serves as input for the networkx function which creates the graph object.

number of gaze transitions across all OOIs. An example of an input data frame is shown in Figure C.5 (Ref. to OSF: `1-2_GazeBasedAttentionNetworks/fl2_transition.py`). As a result, one graph represented one gaze-based attention network for one participant during one experimental session. The networkx graph objects (<graph_name>.p) were visualized with function family around <<networkx.draw()>> and used for further analysis. (Ref. to OSF: `1-2_GazeBasedAttentionNetworks/main.py`)

Tips for application: The gaze transition dataset was grouped by the source and the target variable and counted how often each combination occurred. Proceeding like this for the whole dataset, the number of transitions from each object to all others could be counted, using only one line of code. The complexity of the resulting graphs depended on the number of total OOIs (nodes) and the frequency of gaze transitions (edge weights). The size of the files was reduced significantly by saving networkx graph objects instead of CSV datasets. All transition datasets for our participants had an average size of 40KB, while the stored graph files only had an average size of 1.3 KB.

III. Computing structural variables to describe gaze-based attention networks

To compare the network structures between participants, various measures can be computed to compare values between graphs (e.g. [352], [353]). Our selection of structural variables allowed networks to be compared statistically. All structural variables below can be assigned to one of these categories:

- Centrality measures
- Distribution measures
- Interconnectedness measures.

Centrality measures describe a node's importance or prominence within a network by assessing the number or weight of connections it holds with other nodes [354]. These calculations can be used to determine the importance of certain OOIs in relation to visual attention in the virtual environment. Distribution measures act as a proxy for understanding how visual attention is distributed among specific nodes compared to all others [183]. They provide a means to analyze the distribution of visual attention within a network. Interconnectedness measures focus on the connections between nodes in a network [355]. In gaze-based attention networks, they help to understand how OOIs are linked, i.e., which OOIs build subgroups with frequent gaze transitions. The example code on how to calculate the structural variables is provided in the repository (Ref. to OSF: `1-2_GazeBasedAttentionNetworks/fl3_calculate_graph_features.py`).

Centrality Measure

Degree centrality is a measure calculated as the sum of the weights of a node's incoming and outgoing edges. Previous studies have used this centrality measure to investigate visual attention [145]. It indicates the frequency with which a participant transitions toward a specific object. It is also possible to sum up the degree centrality for a group of OOIs. For comparing its value between participants (i.e., between graphs), one must ensure that one always considers the same group of nodes. To calculate degree centrality, the <<degree>> function of the network package was used. Moreover, its functionality was extended in the code to calculate degree centrality for groups. An example of an undirected network can be seen in Figure C.6a, and the calculated degree of centrality is shown in Figure C.6b.

Distribution Measure

Weighted degree centrality is a distribution measure implemented according to Candeloro et al. [183]. The given formulas were implemented in Python, by changing some aspects, and adding some details. Weighted degree centrality (WDC) can only be calculated for one node and is a measure of the uniformity of all outgoing edges from that node. To calculate WDC, one needs to know the number of outgoing edges (DC), which can be computed using <<Graph.out_edges(node)>>. Given the equations from the paper [183], the formula can be simplified in the following way:

$$WDC = DC \cdot \frac{AUC_{F_c}}{AUC_{max}} = DC \cdot \frac{AUC_{F_c}}{\frac{DC}{2}} = 2 \cdot AUC_{F_c} = 1 + 2 \cdot \sum_{i=1}^{DC-1} Fc(i) \tag{C.4}$$

given that $AUC_{F_c} = \frac{1}{2} + \sum_{i=1}^{DC-1} Fc(i)$. The term $Fc(i)$ in Eq. C.4 is defined as a sum of edge weights. Given the definition in [183] it was calculated as

$$Fc(i) = \sum_{J=1}^{i} \frac{W_J}{\sum_{k=1}^{DC} w_k} \tag{C.5}$$

with w_J being the edge weight for an edge J.

Tips for application: When calculating WDCs for different participants but the same node described in Eq. C.4 and C.5, the outgoing edges needed to be sorted by their weight size to compare uniform distributions between participants. This is not explicitly mentioned in the paper but becomes relevant if the first edge is not always the largest and there are different outgoing edges in different networks. So, before calculating WDC given the formula, the list of edge weights was sorted, $\left[w_J, J \in [1,..,DC]\right].sort()$. For an illustration, see Figure C.6c.

The uniformity measure is another distribution measure implemented using a chi-square test from the scipy.stats Python package [356]. This test for categorical data tests against the null hypothesis that the data is uniformly distributed (when using default arguments). To get a uniformity measure of gaze transitions, the chi-square test statistic can be calculated for all edge weights of a graph, multiplied by a negative one:

$$U = (-1) \cdot scipy.stats.chisquare(\text{list of edge weights}). \tag{C.6}$$

As a result, the higher the U in Eq. C.6, the more the gaze is uniformly distributed across the OOIs. A less uniformly distributed gaze network implies that participants often transitioned between a smaller subset of OOIs while ignoring other OOIs. Our analysis found that a smaller U was also correlated with longer fixation duration on frequently visited OOIs. However, the fact that some nodes are less frequently visited does not necessarily imply a smaller fixation duration for these OOIs. Participants could also focus on single nodes for a long time without transitioning much. For an illustration, see Figure C.6d.

Interconnectedness Measure

The **cut size** is an interconnectedness measure and is especially interesting when dividing the virtual space into different areas. The nodes of the graph can be separated into two unique groups. Cut size calculates the sum of all edge weights between these two groups.

Cut size solely focuses on the connection between the two groups [185]. As a result, a larger cut size indicates more gaze transitions between the two groups, while a smaller cut size indicates less back and forth between the groups. An example can be seen in Figure C.6e.

Tips for application: One can either compute the cut size <<network.cut_size()>> or a normalized cut size with <<network.normalized_cut_size()>>, which is normalized to the sum of the total edge weights. The measures produce different results when comparing the gaze networks of different participants. When asked, "How often do participants transition between two sets of OOIs?" one should compute the cut size. In contrast, if the question is "Compared to all other gaze transitions in the environment, how much do participants transition between two sets of OOIs?" one should calculate the normalized cut size.

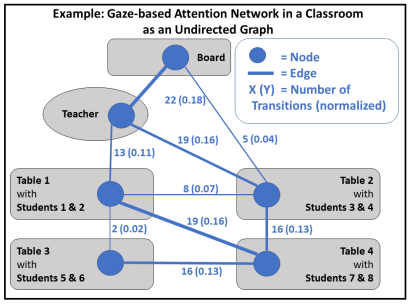
By computing cliques, other structural variables describing measures of interconnectedness could be calculated. A clique is the maximal subset of nodes, where all notes have edge weights larger than zero. This means at least one gaze transition must exist between all subset nodes. A node can be a part of different cliques, but different cliques must have at least one different (less/more/other) node.

The total number of cliques in one network reflects whether participants frequently transitioned with their gaze between objects. A higher number of cliques could be associated with less focused gaze behavior. The average clique size is another valuable variable since it captures the interconnectedness of information gained from cliques. Let us assume that one participant has a higher total number of cliques than another. One could not assume higher interconnectedness if one did not check for the average clique size because one large clique could collapse into two smaller cliques if, for example, one gaze transition was missing. An illustration of the structural variables calculated from cliques can be seen in Figure C.6f.

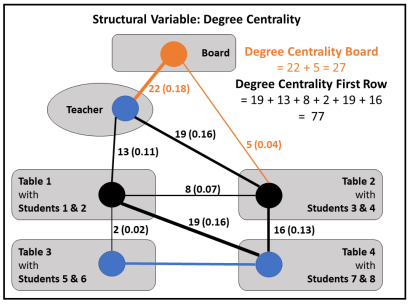
Tips for application: Cliques can only be computed for undirected graphs. Therefore, if the gaze-based attention network consists of directed gaze transitions, one must add respective incoming and outgoing notes and transform the graph into an undirected graph (UG). The networkx package has an implemented function <<network.find_cliques(UG)>>, which can be applied to UGs.

Performance Evaluation

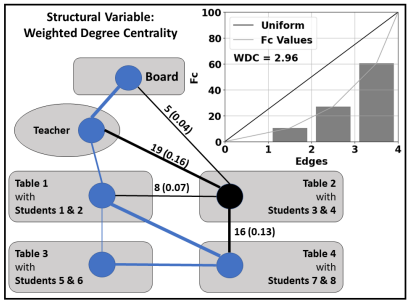
To make this work more applicable to researchers, suggestions for efficient data processing were provided, and the performance of our data pipeline was evaluated. Data aggregation and network analysis, presented in Section C.1.2, were evaluated based on the runtime metric. Specifically, to reach out to social scientists, who often use the programming language R



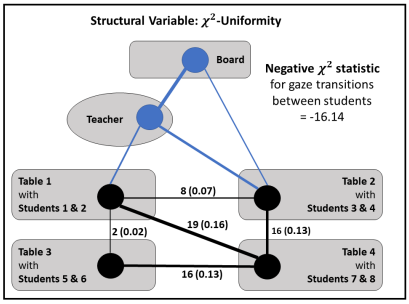
(a) Example network with OOIs as nodes and (normalized) number of transitions displayed as edge weights.



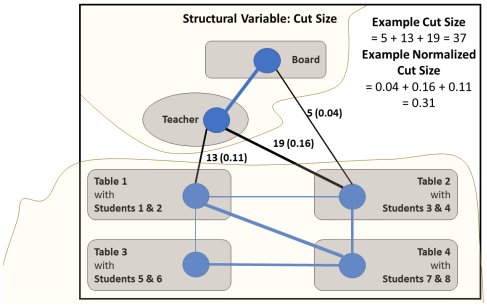
(b) Calculation of degree centrality for either a single node (board) or for a subset of nodes (first row of students).



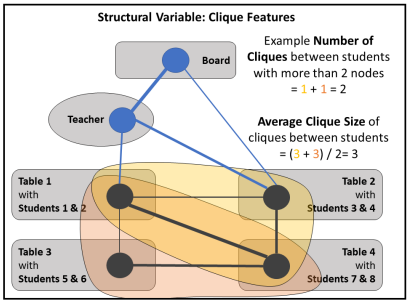
(c) Calculation of the weighted degree centrality for one node (table 2). The top right shows the Fc calculated.



(d) Chi-square uniformity measure, calculated for a subset of all students. Not that this value is always negative due to the formula.



(e) An example of cut size when separating two subsets, namely all students and all non-students (teacher, board).



(f) Computation of all cliques of students larger than 2. The total number of cliques and average clique size can be calculated.

Figure C.6.: Examples of computing structural variables from an undirected graph. A scenario of gaze transitions in a classroom is shown with reduced complexity (fewer nodes) to create a gaze-based attention network for a participant. The example network has the same nodes and edges in all structural variable calculations. A larger display of the example images can be found in the Supplementary Material

instead of Python, the potential advantages and disadvantages of using R (Version 4.3.2) were evaluated. The tests and evaluations, as well as parts of the code programmed in R, can be found in our OSF repository (Ref. to OSF: 1-3_PerformanceEvaluation).

The eye-tracking datasets collected during the VR session differ in length, depending on the experiment duration. Additionally, the available hardware and the rendering complexity of the VR environment influence the framerate of the VR device and, therefore, determine the number of data points collected per second. For our experiment and data analysis, a Lenovo Legion 7 Laptop with an Intel Core i9-10980HK CPU @ 2.40GHz, 3096 MHz, 32GB of RAM, and an NVIDIA GeForce RTX 2080 Super graphic card was used. One data point was obtained on average every 25 milliseconds (40 frames per second (FPS)). This resulted in an average of 36000 data points for the 15-minute VR experiment for one participant (one minute ~ 2400 data points). Note that remote eye trackers provide a higher temporal resolution (e.g., 1000 FPS), which cannot be compared to VR eye tracking and requires different processing and data handling.

The obtained dataset can be processed separately for each participant, so the total runtime would be the runtime of processing a single dataset times the number of participants. Similarly, the data pipeline saves newly aggregated, smaller data sets for each processing step, so the estimated runtime is the sum of all individual processing steps. The runtime for one representative dataset of one participant from the original sample was evaluated (dataset size: 38005 rows). All runtimes of the following steps can be found in Table C.3.

Preprocessing

Working with large datasets, particularly in the context of eye-tracking data, can present challenges such as increased processing times, risk of stack overflow, and overall difficulties in data management. For researchers, especially those new to handling extensive eye-tracking datasets, performance optimization recommendations are offered in the subsequent steps (i.e., preprocessing, dimension reduction, and creating data duplications to offload tasks from primary memory storage (RAM) to secondary memory storage (CPU)) [357].

While some processing pipelines working with large datasets could encounter potential runtime errors [360], the design of our methods is specifically friendly to processing larger datasets. This is because the most memory-intensive task during data processing is parsing the raw data in the first step, while all subsequent steps profit from data reduction. Moreover, the network approach presents itself as a method to reduce data complexity and dimensionality [357], which was identified as a powerful tool in social sciences [361]. In the first step, it

	Clean & create a smaller dataset	Create transition dataset	Create graph objects	Graph Features	
				Number of cliques	Weighted degree centrality
Python	0.58s	0.31s	0.01s	$>0.01s$	$>0.01s$
R	2.56s	0.51s	0.12s	0.01s	0.02s
Potential runtime errors	Loading large datasets can cause potential memory-related errors [357].	Filtering and processing data could encounter errors (missing values or incorrect conditions specified) [358].	Inconsistencies or unexpected data formats in the input data; dense or highly interconnected nodes [359].	Complex connectivity patterns or dense graphs may result in longer execution times or stack overflow errors [185].	
Solutions to runtime encounters	Saving and loading CSV files (and monitoring RAM usage).	Drop missing values and identify placeholders before processing; specify variable type.	Creating graphs from datasets with specified variables; dimensionality reduction by selecting or merging OOIs.	Avoid graph features that need extensive traversing through the graph. Keep the graph size low (see previous point).	

Table C.3.: Evaluated runtime of all processing steps of the data pipeline stated in seconds. Time is measured for one eye-tracking dataset (one participant). Potential runtime errors and how the data pipeline (method) avoids these are stated.

is advisable to narrow down the dataset to a smaller subset that contains only the essential variables required for your analysis (dimension reduction). Typically, datasets from VR experiments may include up to 90 different variables. However, for the methodology discussed in this paper, only two variables are essential (or five if pupil diameters are included for data cleaning). As described in Section C.1.2 (I), the two essential variables are the time and the gaze target variable that contains the names of the OOIs collected via the ray-casting method. If pupil diameter is used for data cleaning, both left and right pupil diameter and, if available, eye-openness variables should be contained in the dataset for the first preprocessing step. Dropping unnecessary variables can lead to a reduction in dataset size by approximately 95% -for instance, reducing the dataset size from around 33 Megabytes (MB) to just 1.7 MB.

Additionally, these datasets often feature missing data or include OOIs irrelevant to the study. Eliminating these elements during the initial preprocessing steps results in more manageable dataset sizes and enhances processing efficiency in subsequent stages (preprocessing). In our experience, the time required for cleaning and saving a dataset for a single participant was 0.58 seconds in Python and 2.56 seconds in R. Additionally, processing each dataset separately and freeing memory after saving the data creates data duplications that offload task complexity from primary memory storage (RAM) to secondary memory storage (CPU) and distributes resources more evenly across hardware memory [362]. By adopting these strategies, researchers can significantly mitigate the computational challenges associated with large datasets and avoid potential runtime errors (Ref. to Table C.3).

I. Aggregating raw gaze-target information into gaze transition datasets

For the preprocessed, smaller datasets, the initial step involves generating the transition datasets through a process that iterates over all rows in the dataset via a single 'for loop.' Consequently, the runtime of this operation exhibits a linear relationship with the size of the dataset (denoted as N), leading to a computational complexity of O(N) according to Big O notation. In terms of performance, this process took 0.31 seconds in Python and 0.51 seconds in R.

II. Creating gaze-based attention networks from gaze transition datasets

In the Python pipeline, an additional step was introduced to generate adjacency matrices compatible with the networkx package. This allowed the pipeline to construct graph objects directly from the data. Similarly, in R, the adjacency matrices were computed first, and these matrices were then used to create graph objects with the igraph library. The execution time for

this process was 0.01 seconds in Python and 0.12 seconds in R. It's important to note that the dataset sizes progressively reduced through aggregation steps, resulting in a final graph that only includes OOIs from the environment as nodes, with the number of transitions between them as edges. In the performance evaluation, we deliberately chose not to filter out any OOIs, leading to a graph with 91 nodes, reflecting the 91 distinct gaze targets identified in the VR environment. This contrasts with the 26 nodes used in our original analysis. The number of OOIs, inherently constrained by the VR environment, thus determines the maximum size of our graphs.

III. Computing structural variables to describe gaze-based attention networks

Given the initial low runtime for graph creation, the runtime was only measured for two structurally complex variables: the calculation of clique numbers and weighted degree centrality. Consequently, R code was added to the repository for these calculations. The computation time for the clique numbers was less than 0.01 seconds in Python and 0.01 seconds in R. For weighted degree centrality, Python completed the task in less than 0.01 seconds, whereas R took 0.02 seconds. The performance of other structural variables, such as degree centrality, was not assessed because functions for these calculations are readily available in libraries like networkx (e.g., graph.degree()) and igraph (e.g., strength(graph)).

The analysis revealed no significant performance issues in either Python or R, with R consistently showing slightly longer runtimes for all tasks [363]. Despite the requirement for high-performance hardware to run VR experiments using the HTC VIVE, the data analysis procedures only necessitate the computational capabilities of standard hardware. This aspect underscores the efficiency of our methodology. While calculating eye movement features requires several iterations over the entire data set, this method allows the data to be condensed swiftly and efficiently into more manageable formats. The most performance-intensive aspect of our analysis is the initial data cleaning and reduction process.

Considerations for Implementation and Application

For the method presented, some aspects should be considered for implementation. One important aspect is that the accuracy and precision of the integrated eye tracker affect the gaze-ray casting technique [58]. While the ray-casting technique collects data every time stamp, the measured information could be of varying quality. For example, if the OOIs in the environment are too small, the global gaze vector might miss the focused object. This can later be seen in the data when, for successive time stamps, different objects are tracked

alternately. To avoid additional processing steps for data cleaning, the virtual environment should consist of larger OOIs. If this is not the case, one must consider merging smaller OOIs into bigger ones. The necessary size of an OOI can be determined by performing a test run before the experiment, where a test person should be asked to look at many smaller objects in the environment.

Another important aspect concerns the number of OOIs that are used in a network. The comparison of networks between participants by analyzing structural variables is not affected by the number of OOIs added to the network. Using a large number of different OOIs for the analysis might only influence the clarity of visualization. In contrast, when merging OOIs, the resulting networks might be under-complex. Imagine a network with only two large OOIs that is created by merging many smaller OOIs. Analyzing cliques would not be possible here because there would only be one clique, the trivial one. Moreover, two limitations must be formulated when using the method. First, it is important to note that our method only focuses on overt visual attention and does not cover peripheral perception (covert attention) or recognition aspects. A second limitation is that our method was only evaluated using a relatively static environment, where the OOIs did not move too much. The analysis of gaze-based attention networks in VR with moving objects might be more tentative. We recommend that future applications using our method investigate more dynamic virtual environments with moving objects. While analyzing moving scenes usually requires frame-by-frame object detection, our approach already incorporates this by using the physical objects as OOIs. Since the OOI position can be extracted from the gaze-ray casting pipeline, this information can be used to aggregate node feature information and be further processed.

Furthermore, the structural variables described in this article are only a selection of the measures that can be computed from graphs. Other measures could be considered to analyze gaze transitions. While degree centrality translates directly into a measure of attention distribution towards (groups of) OOIs, other centrality markers are closeness centrality, between centrality, or eigenvector centrality [184]. Depending on their calculation formula, these measures require different interpretations when applied to gaze-based attention networks. Most calculated variables can be used to perform statistical analyses. However, one must be aware that the values of the structural variables could be non-normally distributed.

Collecting node feature information can also extend our statistical analysis of structural variables using graphical neural networks (GNN) models. From our provided pipeline, the aggregated networks can be directly transformed into graph representations for GNNs with the pytorch.geometric [364] library. Some structural variables (like degree centrality or clique

information) can even be incorporated as node features in the GNN. The presented network analysis can also be performed similarly with the R programming language (R Core Team, 2022). Since R also has a data frame object type, the data can be transformed into graphs using, for example, the igraph package [365], which allows for the calculation of structural variables. Especially for statistical analysis and visualization, R might be a suitable choice. In contrast, state-of-the-art machine learning implementations using networks and graphs are provided in Python.

Our large empirical evaluation with $N = 274$ students showed that the method can be successfully applied and has proven that structural variables show great potential for analyzing students' learning and social behavior in a VR environment. Future work should consider exploring the changes over time to examine the dynamics of gaze behavior within the given VR experience. The provided data structure also allows for modeling temporal graphs [187], [366], [367], which could be adapted in future research. Moreover, the full potential of this analysis might be revealed when applied to different VR settings and tasks, like visual exploration, navigation, or joint attention.

C.1.3. Ethics statements

In this study, all procedures involving human participants were in accordance with the ethical standards of institutional and/or national research committees. Ethical approval for this study was obtained from the university's Ethics Committee prior to the beginning of the research. The Ethics Committee reviewed and approved all aspects of the study to ensure that they adhered to ethical principles and standards. The study was conducted in accordance with the ethical guidelines set forth by the Declaration of Helsinki, as well as other relevant regulations and laws governing research involving human subjects. Participants were informed about the purpose and procedures of the study and provided written consent prior to their participation. All data collected during the study were treated with strict confidentiality and anonymity to ensure that the privacy and well-being of participants were protected. Any identifiable information was removed or pseudonymized before the analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data is shared on https://osf.io/pek4q/?view_only=ef151fd06ac8413a827020d4264b3c8d as part of the co-submission.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mex.2024.102662.

C.1.4. Acknowledgments

This research was supported by a grant to Richard Göllner funded by the Ministry of Science, Research and the Arts of the state of Baden Württemberg and the University of Tübingen as part of the Promotion Program for Junior Researchers. Philipp Stark is a doctoral student at the LEAD Graduate School & Research Network, which is funded by the Ministry of Science, Research and the Arts of the state of Baden-Württemberg within the framework of the sustainability funding for projects from Excellence Initiative II. We want to thank Jens-Uwe Hahn, Stephan Soller, Sandra Hahn, and Sophie Fink from the Institute for Games, Department of Computer Science and Media at the Hochschule der Medien Stuttgart for their extensive work preparing the immersive virtual reality classroom used in this study. We acknowledge support from the Open Access Publication Fund of the University of Tübingen.

C.2. Learning with simulated virtual classmates: Effects of social-related configurations on students' visual attention and learning experiences in an immersive virtual reality classroom

C.2.1. Abstract

Immersive virtual reality (IVR) provides great potential to experimentally investigate effects of peers on student learning in class and to strategically deploy virtual peer learners to improve learning. The present study examined how three social-related classroom configurations (i.e., students' position in the classroom, visualization style of virtual avatars, and virtual classmates' performance-related behavior) affect students' visual attention toward information presented in the IVR classroom using a large-scale eye-tracking data set of N = 274 sixth graders. ANOVA results showed that the IVR configurations were systematically associated with differences in learners' visual attention on classmates or the instructional content and their overall gaze distribution in the IVR classroom (Cohen's d ranging from 0.28 to 2.04 for different IVR configurations and gaze features). Gaze-based attention on classmates was negatively related to students' interest in the IVR lesson (d = 0.28); specifically, the more boys were among the observed peers, the lower students' situational self-concept (d = 0.24). In turn, gaze-based attention on the instructional content was positively related to students' performance after the IVR lesson (d = 0.26). Implications for the future use of IVR classrooms in educational research and practice are discussed.

Keywords: immersive virtual reality, classroom simulation, peer effects, visual attention, network analysis, eye-tracking

C.2.2. Introduction

Searching Web of Science for peer-reviewed articles with "virtual reality" and "education" in the Abstract yielded about 3,600 results—two-thirds of which were published within the last 5 years (according to a search as of October 2021 using the Web of Science database and searching for peer-reviewed articles including the keywords virtual reality AND education in the Abstract). From immersive virtual reality (IVR) applications for engineering education [296], the military [368] and medical training [295], [369] to environmental education [370], virtual field trips and science simulations in elementary and secondary school [101], [102],

[298], [371]: IVR and its associated affordances are becoming more and more popular in training and education. Most educational IVR applications focus on experiential learning, particularly simulations of experiences that are difficult or impossible for learners to have in real life [28], [372]. However, in addition to IVR simulations that take learners out of the classroom, the transformation of "typical" classrooms into IVR learning environments is a promising methodology for educational research and practice. Understanding learning environments as all sorts of surroundings in which learning can take place, the present study focuses on regular classroom settings as a specific learning environment with certain structural elements (e.g., a blackboard/screen, a class of peer learners, seating at tables in rows). These structural elements contextualize learning for students in the classroom and can be transferred to and utilized in IVR learning environments. IVR technology makes it possible to create computer-generated simulated environments that allow for realistic perceptions and seemingly real interactions within an artificial and hence fully controllable virtual world (e.g., [54]). An IVR classroom thus provides a simulated classroom environment that learners experience in a manner that is similar to how they experience a classroom in the real world; however, the IVR classroom and included virtual characters can simultaneously be freely designed and fully controlled with regard to their appearance, behavior, and interactions (e.g., transformed social interactions; [91]). IVR classrooms thus allow educational researchers and practitioners (a) to examine the relevance of different classroom features for student learning and (b) to strategically deploy these features to design IVR classroom environments that further enhance the potential of traditional classroom learning. In this vein, IVR technology in general—and as we argue, IVR classrooms in particular—can lead to promising "varied educational contexts that have the potential to enhance (and, thus, alter) the process of learning in significant ways" as Alexander [373] phrased it in her treatise on research in education psychology (p. 156). Importantly, in order to use IVR classrooms to move beyond what has been gleaned from centuries of traditional classroom learning and research, there is a need for systematic studies that (a) are based on established learning theories and simultaneously (b) provide insights into exactly how the potentials of new technologies can be exploited to improve and further enhance classroom teaching and learning (e.g., in remote learning scenarios). The present study aims to provide initial systematic insights into how IVR classrooms can be used to advance educational classroom research as well as learning and teaching practices, especially with regard to the central role of peer learners.

Notably, classroom situations in the real world are complex and dynamic, and students' classroom learning is substantially shaped by numerous contextual and peer-related factors

[374]–[376]. The (perceived) learning environment—which is strongly characterized by peer learners—has been found to be related to students' achievement and academic trajectories [202], [377]–[380] as well as their emotions and motivation during learning [203], [381], [382]. Assuming that peer learners substantially shape students' learning experiences not just in real-world classrooms [383], [384] but also in IVR classroom settings, it is crucial to understand how respective peer effects can be examined and utilized by transforming "traditional" classmates into virtual peer learners (i.e., avatars of either other human participants or simulated computer-based co-learners in the IVR classroom). In the present study, we focus on virtual peer learners as a group of fully preprogrammed social counterparts (i.e., a simulated virtual class) for individual participating students in an IVR classroom. A central goal when designing IVR classrooms for educational research and practice should be to authentically simulate classroom scenarios that include the social counterparts in order to (a) use IVR classrooms as an experimental tool to gain insights into the social-related processes underlying students' learning in the classroom (i.e., in a standardized yet authentic setting; [54], [385] and subsequently to (b) strategically deploy certain social-related configurations for more effective learning in IVR classrooms (e.g., in remote learning scenarios or using virtual peer learners as pedagogical agents; [91], [102], [386]). In the context of peer effects, particularly the perceived proximity versus distance (i.e., also similarity vs. dissimilarity) to peer learners is considered a critical aspect with regard to the effects that social contexts have on students' learning experiences [383], [387], [388].

Whereas the use of IVR classrooms in educational research and practice has been increasing (see examples by [66], [76], [91], [95], [96]), there is a scarcity of systematic insights into how different configurations, specifically in the IVR classroom, affect users' perception of the IVR environment and virtual social counterparts. Importantly, the majority of existing studies about individual IVR experiences are based on samples of (young) adults; hence, a clear understanding of how children perceive IVR environments and social interactions in the virtual space is lacking [328]. Aiming to address this issue, the present study focuses on the following question: How exactly do different configurations of social-related features in an IVR classroom affect how and the extent to which students attend to the (social) information provided during an IVR lesson? We thereby particularly focus on social-related configurations regarding learners' proximity to their simulated virtual social counterparts (i.e., spatial, visual, behavioral aspects of proximity).

Of course, there are countless ways to configure an IVR classroom and therefore many features that could potentially influence students' attention toward (social) information in

the IVR environment. However, some configuration features are more salient and socially relevant than others, such as the perspective from which students experience the IVR classroom and virtual peer learners (i.e., spatial proximity) or the visualization style and behavior of their social counterparts (i.e., visual and behavioral proximity). Do students focus more on their virtual classmates versus the instructional content when they sit in the back of the IVR classroom? What role does a more or less stylized visualization of virtual classmates play? Finally, does virtual peer learners' performance-related behavior (e.g., more or less hand-raising) affect students' learning experiences in the IVR classroom?

In the present study, we aim to provide answers to these questions by examining students' learning experiences in an IVR classroom with different configurations. More specifically, we examined three social-related features of IVR classroom configurations that are decisive for how students attend to what is happening during a virtual classroom lesson (see Section C.2.2). To gain insights into students' learning experiences in the IVR classroom, we used students' gaze data and analyzed their gaze-based attention networks in the different IVR configurations (see Section C.2.2). In order to provide insights into the meaning of the gaze-based features used, we additionally examined how they are related to central learning outcomes, namely students' interest in the lesson, their situational self-concept and post-lesson achievement.

Configuration of immersive virtual reality classrooms for educational research and practice

Given the myriad of decisions involved in the configuration of IVR classrooms, findings from educational psychology research and already existing studies in IVR (classroom) contexts point to central social-related features that seem to affect students' learning experiences in the classroom and therefore need to be carefully considered when configuring IVR classrooms. We focus on the following configuration features that represent variations in spatial, visual, and behavioral proximity of the virtual social counterparts, respectively:

(1) students' position in the IVR classroom (i.e., the view of virtual classmates and the instructional content from a front or back row)

(2) the visualization style of social counterparts (i.e., the style of the avatars used to represent simulated virtual classmates and the virtual teacher)

(3) virtual classmates' behavior and performance.

First, one of the most salient and socially relevant features for students' classroom learning is the seating arrangement and position of students within the classroom [389], [390]. Research on the effect of students' position in the classroom has provided mixed findings regarding different outcome variables, indicating a positive effect of a front seating position close to the teacher on students' performance, but also null effects for performance outcomes or only effects on students' motivation, not performance [391]–[398]. This is not surprising considering that due to natural limitations in the classroom, existing studies are situated in very different classroom environments and have not always randomly allocated students to seating positions (i.e., results that indicate better learning outcomes for students sitting in the front might be confounded by the seating choices of higher-performing students and/or changes in teachers' instructional practices in response to certain classroom compositions). In order to address this issue, some IVR studies have systematically varied students' position in the classroom in order to provide experimental evidence. For instance, Bailenson et al. [91] manipulated participants' position in the classroom, including their distance to the teacher, while keeping all other factors constant and found effects on students' subsequent learning outcomes (see Experiments 2 and 3). Similarly, Blume et al. [66] found that students who were assigned to a position closer to the virtual teacher performed better in a posttest compared to students that were placed in the back of the IVR classroom. Most importantly for the scope of the present study, existing research has not moved beyond learning outcomes as a measure of distinct effects of seating positions in the classroom on students' learning experiences. Thus, how students' position in the classroom actually affects how they attend to (social) information in the classroom (e.g., the instructional content and social information provided by their classmates) remains an open question. Whereas sitting in the front is most likely associated with increased attention to the teacher and instructional content, paying some attention to peers might also be desirable, particularly when considering potentially beneficial effects of peers (e.g., as pedagogical agents or as a motivating reference group).

Second, another very salient and socially relevant feature when configuring IVR classrooms is the visualization style of social counterparts such as virtual classmates and the virtual teacher [399]. As the Uncanny Valley effect [400] and related works (e.g., [401], [402]) indicate, avatars' more human-like appearance is not the only decisive factor, and more importantly, not always desirable if the goal is for users to have a favorable perception of virtual avatars [403], [404]. Notably, IVR studies examining this effect mostly compare the two ends of the spectrum, i.e., full-body human-like avatars versus non-human-like visualizations such as avatars with only a head and hands [405], a drone that functions as a pedagogical agent

[406] or avatars with a more animal-like appearance [407]. If the aim is to configure an IVR classroom with a teacher and classmates that are clearly recognizable as such and able to simulate a real-world classroom scenario, how realistically these human-like avatars need to be visualized remains an open question. Does a cartoonish visualization of virtual classmates and the virtual teacher lead to the same perceptions as more stylized representations? Particularly given that animation and design costs increase with increasing realism, it seems worth examining what degree of realism is necessary when designing virtual avatars. Moreover, in addition to the question of what is perceived as authentic and realistic—which has been the focus of most avatar-related research to date, another open question concerns how different avatar visualization styles affect students' visual attention toward social information provided in the classroom (e.g., virtual classmates' behavior in contrast to the instructional content). For instance, previous studies have found longer fixation durations in an IVR classroom with cartoon-style avatars [37] and longer dwell time on peer learners visualized as cartoonish characters [3]. When interpreting the results, the authors argued that the unusual appearance of cartoon-style peer learners and the increased difficulty of decoding social information from less realistic avatars might lead to these results.

Third, another socially relevant feature to consider when configuring IVR classroom scenarios is peer learners' behavior (e.g., performance level and active participation). Educational psychology research has repeatedly demonstrated that classmates substantially shape student learning, highlighting the role of what can be called "classroom composition" effects (see, e.g., [384]). On the one hand, there is evidence for so-called positive spillover effects of higher-achieving peers on students' achievement and self-evaluations, in the sense that learners benefit from high-achieving peers and perform better when they are surrounded by high-performing classmates [408]–[411]. On the other hand, there is a large body of evidence for negative contrast effects in the face of high-achieving classmates, suggesting that higher-achieving peers have a negative impact on students' evaluations of their own competence, controlling for individual achievement (the so-called Big-Fish-Little-Pond Effect, see latest reviews by [412]–[414]).

Notably, whereas the aforementioned peer effects have been studied exhaustively in educational psychology research—typically by examining students' test performance and self-reports of their own competencies in relation to their peers' average test performance, the effect of peer learners' actual (performance-related) behavior has received little attention. In other words, the actual processes underlying peer effects in a classroom situation (e.g., the effect of peer learners' behavior on students' learning and attention distribution) remain

largely unexplored. IVR classrooms provide the opportunity to examine such effects in an authentic yet controlled setting, as demonstrated for instance by Bailenson et al. [91]. The authors manipulated virtual classmates' attention-related behavior (i.e., peer learners being attentive or distracting during instruction; see Experiment 4) and found positive effects of more attentive virtual classmates on students' performance after the IVR lesson. Similarly, a study in the field of economics manipulated virtual co-workers' productivity and found a positive relation with the performance of participants doing the same task as the virtual co-workers in an IVR environment [415], [416]. In sum, based on existing IVR studies, it can be assumed that classmates' performance affects students' learning in IVR settings; however, it is still unclear how exactly peers' (performance-related) behavior in an IVR classroom needs to be configured in order to be recognized by K-12 students in an IVR classroom scenario. Beyond the opportunity to use IVR classrooms as a tool to examine the effects of peers' performance-related behavior, such studies have important implications for the design and use of virtual peer learners as pedagogical agents in IVR classroom-based learning applications (see, e.g., [91], [102], [386]).

Taken together, the three outlined features of IVR classrooms (i.e., students' position, visualization style of virtual avatars, and virtual peer learners' behavior) play an important role in student learning. However, with previous studies based either on real-world classroom research or self-reported experiences in IVR (classroom) settings, it remains unclear how exactly these features affect how students attend to different types of information in an IVR classroom environment. Students' gaze data provides an opportunity to obtain such insights.

Students' gaze-based attention networks

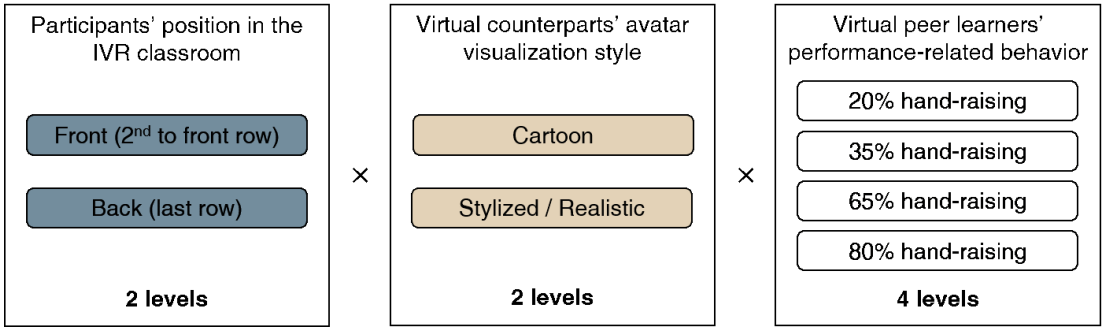
Students' gaze behavior allows for insights into how students attend to information presented to them in an IVR classroom environment [106], [266], [417]. Moreover, compared to real-life classrooms, gaze data from an IVR classroom provides the opportunity to combine the high methodological rigor of a standardized environment with an authentic representation of a classroom situation with all its accompanying dynamics. Thanks to recent technological advances, state-of-the-art IVR equipment comes with integrated eye trackers, making it possible to unobtrusively examine students' gaze behavior to gain an unbiased and in-depth understanding of their learning experiences in the IVR.

Importantly, the interpretation of eye-tracking data is known to be context-specific, and the appropriate analysis technique to understand how attention is distributed must be chosen carefully [61], [62]. Most learning-related studies analyzing eye movement data have used

temporal and count measures such as number of fixations, number of saccades, and fixation durations [62], which are easy to collect with available software. However, these commonly used eye-tracking features are often analyzed in isolation, and it is difficult to establish an interpretable link between eye movements and underlying cognitive processes (e.g., [418]). As Lai et al. [62] point out, more sophisticated measures are necessary to investigate meta-cognitive skills in-depth. Based on the assumption that students guide their attention in the classroom and focus on certain objects (e.g., their classmates) while ignoring others (e.g., the teacher and instructional content on the screen), what students look at can serve as an indication of what they pay attention to. Such so-called overt spatial or visual attention [110], [114], [419], [420] can be analyzed using eye-tracking data. Considering that an IVR classroom is a relatively static environment where spatial relations between objects do not change substantially over time, it can be assumed that students are able to willfully direct their attention to certain objects at least to a substantial degree [421], [422]. Corresponding processes of active information gathering are reflected in longer gaze movement periods, such as consecutive gaze shifts from object to object (instead of eye movement features like fixations and saccades operating on a level of milliseconds; [61]).

Taking these aspects into consideration, in the present study, we opt for a rather novel approach and apply the methodology of network analysis to the analysis of gaze data to gain insights into students' gaze-based attention distribution in an IVR classroom. Network analysis (based on the mathematical theory of graphs; [182]) is a prominent method in various scientific fields, including biology, geography and the social sciences (e.g., [201], [423], [424]). However, this approach has so far received little attention in eye-tracking research and there are only few studies performing network analysis with gaze data [145]–[147], [150]. We argue that particularly when it comes to students' visual attention toward (social) information in a classroom situation, the analysis of gaze-based attention networks provides novel and most importantly explainable and interpretable insights into students' gaze behavior during the IVR experience. It makes it possible to identify the degree to which certain objects of interest (e.g., peer learners, the teacher or the instructional content) are in the center or focus of gaze transitions (via so-called gaze centrality markers, see, e.g., [145]). Moreover, the analysis of gaze-based attention networks allows for extracting information about the overall gaze activity and connectedness of gazes between certain objects of interest (e.g., how intensely different peer learners are attended to) or the overall distribution of gaze between objects of interest in the environment (e.g., how often students' gaze goes back and forth between the teacher and peer learners). A detailed description of the method and corresponding visual

Figure C.7.: 2 × 2 × 4 Between-Subjects Design With Different IVR Classroom Configurations



attention measures can be found in Methods Sections 3.4.1 and 3.6.1 and in Appendix A.

Taken together, analyzing gaze shift movements as gaze-based attention networks has a number of advantages: First, the existing large body of research on graph theory and network analysis provides a rich set of graph properties and variables to describe graph structures, which can be computed from the given gaze-based attention networks. Second, these structural variables are already presented as aggregated variables, which allows for statistical testing and significance analysis without the need for machine learning or complex non-linear regression models. Third and finally, the calculated structural variables can be interpreted on a theoretical level because they directly reflect characteristics of gaze-based visual attention.

C.2.3. The present study: aims and research questions

The present study aims to gain insights into how different configurations of an IVR classroom with a full class of more than 20 simulated virtual peer learners impact students' attention distribution toward (social) information in the IVR classroom and their learning experiences. To extend existing research, we focused particularly on the IVR experiences of children, who have been the subject of considerably less IVR experience-related research to date [328]. To this end, the present study examined how different social-related IVR classroom configuration features (i.e., variations in spatial, visual, and behavioral proximity of the virtual social counterparts) affect students' gaze-based attention networks during instruction in an IVR classroom.

We focused on three socially relevant configuration features that we consider of particular interest when aiming to answer the question of how to examine and utilize peer effects

in an IVR classroom for ideal learning outcomes as well as research purposes, namely (a) participants' positioning in the IVR classroom, (b) the visualization style of virtual avatars of peer learners and the teacher, and (c) virtual peer learners' performance-related behavior. We (a) placed participating students either in a front or the back row of the IVR classroom and (b) visualized virtual avatars either in a cartoonish or more stylized (i.e., more realistic) manner. Moreover, we (c) used peer learners' hand-raising behavior as an indicator of students' behavioral engagement and performance and varied the proportion of virtual classmates who raised their hands to respond to the virtual teachers' question during the IVR lesson (i.e., 20%, 35%, 65% or 80%). Drawing on graph theory, we mapped students' visual attention patterns during the IVR lesson in terms of the gaze allocation to different objects of interest (OOIs; i.e., virtual peer learners, the virtual teacher, and the screen with instructional content) in the form of a graph. We then extracted different features that allowed us to describe students' gaze-based attention networks with regard to the focus of gaze transitions on OOIs, the connectedness of gazes between OOIs and the uniformity of gaze distribution across OOIs in the IVR classroom (see details in Methods Section 3.4.1). We used these features to examine differences in students' gaze-based attention networks with regard to the different IVR configuration conditions, asking:

RQ 1. How do different social-related IVR configurations affect students' gaze-based attention networks in the IVR classroom? More specifically, how do participants' position in the IVR classroom (spatial proximity; front vs. back), the visualization style of virtual avatars (visual proximity; cartoonish vs. stylized) and the performance-related behavior of virtual peers (behavioral proximity; proportion of classmates who raise their hands) affect (a) the degree to which an OOI is in the center/focus of gaze transition, (b) the connectedness of gazes to peers, and (c) the uniformity of gaze distribution across OOIs and in the IVR classroom in general? We used different structural features to assess (a) to (c) respectively and tested the following hypotheses:

H1a. Given that students positioned in the back row of the classroom had the whole class of virtual peer learners in front of them, while students who were positioned in the front had only one row of students between themselves and the teacher and screen, we expected being positioned in the back of the virtual classroom leads to more gaze centrality on virtual peer learners (and less centered gaze networks on the virtual teacher and screen), more connectedness of gazes among peers, and due to the increased field of view, a more uniformly distributed gaze in the IVR classroom.

H1b. Based on existing findings regarding the less usual appearance of cartoonish peer

learners and the increased difficulty of decoding social information from them due to their less fine-grained visualization [3], [37], we hypothesized that a cartoonish visualization of avatars leads to more gaze centrality on virtual peer learners and less gaze centrality on the virtual teacher (gaze centrality on the screen not affected), more connectedness of gazes among peers, and due to the increased focus on peer learners, a less uniformly distributed gaze overall in the IVR classroom.

H1c. Assuming that increased activity of virtual peer learners attracts more attention from students, we expected that more hand-raising behavior of virtual classmates leads to more gaze centrality on peer learners (and gaze networks less centered on the teacher and screen), more connected gazes among peers, and based on the desire to obtain a comprehensive picture of peer learners' behavior, a more uniformly distributed gaze in the IVR classroom.

H1d. We expected the effects of avatar visualization style (H1b) and the variation in peers' performance-related behavior (H1c) on students' gaze-based attention networks to be particularly pronounced when sitting in the back, where more peer learners were in the field of view. We therefore explored interaction effects between the configuration conditions.

In the present study, we analyzed eye-tracking data from an IVR classroom situation with the methodological approach of network analysis. To obtain a more substantiated understanding of how the resulting indicators of students' gaze behavior were related to students' learning experiences in the IVR classroom, we examined relationships with learning-related outcomes. Therefore, in a second step, we asked:

RQ 2. How do structural features of students' gaze-based attention networks (i.e., the degree to which an OOI is in the center/focus of gaze transition, the connectedness of the gaze networks among peers, and the uniformity of gaze distribution across OOIs and in the IVR classroom in general) relate to their learning experiences in the IVR classroom? Students' learning experiences in the IVR classroom were examined in terms of (a) their interest in the IVR lesson, (b) their evaluation of their own competence in the IVR lesson (i.e., situational self-concept), and (c) their performance on a posttest assessing the IVR lesson content. We examined the following exploratory hypotheses:

H2a. We expected that more gaze centrality on peers and a higher connectedness of gazes among peers (i.e., more visual attention on social information) are related to lower interest in the IVR lesson, lower situational self-concept and lower test performance after the IVR lesson.

H2b. Based on the assumption that increased focus of visual attention on the instructional content is beneficial for students' learning outcomes, we hypothesized that more

Table C.4.: Descriptive Sample Statistics after Randomization to One of the IVR Configuration Conditions.

Variable	Total	Front (N = 122)		Back (N = 152)	
		Cartoonish (N = 56)	Stylized (N = 66)	Cartoonish (N = 94)	Stylized (N = 58)
Age	11.50 (0.55)	11.57 (0.57)	11.49 (0.53)	11.47 (0.58)	11.52 (0.50)
Gender					
Female	138	23	36	43	36
Male	136	33	30	51	22
Grades a					
Math	2.61 (0.89)	2.72 (0.87)	2.57 (0.77)	2.51 (0.96)	2.74 (0.91)
German	2.48 (0.72)	2.50 (0.68)	2.51 (0.83)	2.40 (0.72)	2.57 (0.62)
Prior IVR experience b					
No	114	21	25	38	30
Yes	156	35	41	54	26
n/a	4	-	-	2	2
General self-concept intelligence c	3.07 (0.61)	3.20 (0.61)	2.96 (0.56)	3.12 (0.64)	3.00 (0.60)
Initial CT interest c	3.14 (0.76)	3.27 (0.63)	3.18 (0.75)	3.10 (0.82)	3.04 (0.77)

Note. Mean values and standard deviations M (SD) are shown for continuous variables, categorical variables are shown in absolute numbers. Values are averaged across hand-raising conditions. CT = Computational Thinking; IVR = Immersive Virtual Reality. a Grades were on a scale from 1–6 with lower numbers indicating better achievement; b Prior IVR experience was assessed via one item asking whether participants had previously used IVR glasses; c Measured on a 4-point rating scale with higher values indicating higher levels of the respective variable.

gaze centrality on the teacher and screen (i.e., more focus on the instructional content and less attention on social information) are related to higher interest in the IVR lesson, higher situational self-concept and better test performance after the IVR lesson.

Moreover, we explored how uniformity of gaze distribution across OOIs in the IVR classroom (as an indicator of rather balanced attention toward social information) is related to students' interest in the IVR lesson, situational self-concept and test performance after the IVR lesson.

C.2.4. Method

The present study was approved by the regional educational authorities and the ethics committee of the University of [Institution blinded for review] who confirmed that the procedures were in line with ethical standards for research on human subjects (date of approval: 11/25/2019, file number: A2.5.4-106_aa).

Research design

This study follows a 2 × 2 × 4 between-subjects design in which we examined three different IVR configuration features, namely, (a) participants' positioning in the IVR classroom, (b) the

visualization style of virtual avatars, and (c) the performance-related behavior of virtual peer learners (see Figure C.7).

Participants' position in the IVR classroom and virtual avatar visualization were varied on two levels (front vs. back and cartoon vs. stylized, respectively). Virtual peer learners' performance-related behavior was manipulated on four levels via varying proportions of students raising their hands in response to questions from the virtual teacher. A more detailed description of the IVR configurations is provided in Treatment and Materials Section 3.3.1. Participating students were randomly assigned to one of the 16 ($2 \times 2 \times 4$) IVR configuration conditions via random number generation at the individual level. Table 1 shows the descriptive sample statistics after randomization.

Assuming small- to medium-sized effects ($f = .20$), we computed an a priori power analysis for respective analyses of variance with two-tailed tests at a 0.05 alpha level and a minimum power of .90. Based on this, a necessary sample size of $N = 22$ students per group was determined.

Population and sample

We collected data from a total of $N = 381$ sixth-grade students attending academic-track schools in southern Germany. In this study, we used data from $N = 274$ students with a sufficiently high eye-tracking ratio (> 90%). The lack of suitable eye-tracking data from the excluded students was mostly caused by hardware-related problems during data collection (e.g., incorrect eye-tracker calibration, unexpected crashing and restart of the IVR experience) and synchronization issues during data pre-processing. Importantly, the availability of suitable eye-tracking data was unsystematic with regard to the different testing groups and central sample characteristics (see respective statistics in the supplemental material). Similarly to the full sample, the students in our study stem from a total of 25 sixth-grade classes at 14 academic-track schools ($M_{Age} = 11.50$, $SD_{Age} = 0.55$, 50.4% girls). None of the children in our sample had participated in any previous IVR studies, but 57.8% indicated that they had experienced an IVR environment as a consumer at least once before.

Treatment and materials

IVR classroom configuration conditions

We implemented different IVR configurations with regard to the three socially relevant features we consider of particular importance when designing an IVR classroom for ideal learning and research outcomes (see Figure C.7).

We varied the *positioning of participating students in the IVR classroom*, placing them either in a front or a back row. This made participants experience the IVR lesson either (a) from a position close to the instructional center, with only one row of students between themselves and the teacher and screen on which the lesson content was presented, or (b) from a position in the back row of the classroom with the whole class of peer learners between themselves and the teacher and screen (see Figure C.8).

Figure C.8.: IVR Configuration Conditions

Note. The images show different configuration conditions (between-subject variation) taken from the same hand-raising situation in the first phase of the IVR lesson (same instructional content across conditions). The top image shows the avatar visualization in cartoon style; the bottom image depicts the more stylized (i.e., more realistic) avatar visualization. In the top image, 20% of peer learners raise their hand, compared to 80% in the bottom image. The numbers in black circles indicate the seating positions (1) in the front (i.e., the second row) and (2) in the back (i.e., the last of four rows). All virtual characters in the classroom (i.e., the teacher and all peer learners) are simulated and fully preprogrammed avatars.

Moreover, we varied the *visualization style of the virtual avatars* (i.e., teacher and peer learners). Participants were either surrounded by cartoonish or stylized (i.e., more realistically visualized) virtual avatars (see Figures C.8 and C.10 for an impression). The visualization

style only concerned the look of the virtual avatars (i.e., tinier arms and legs, larger heads and eyes, and less fine-grained facial expressions for the cartoonish avatars), whereas audio and motions were the same in all conditions.

Lastly, we varied the *performance-related behavior of virtual peer learners* via their hand-raising behavior; whenever the virtual teacher asked a question during the IVR lesson, either 20%, 35%, 65% or 80% of the virtual peer learners raised their hands to indicate that they knew the correct answer (see Figure C.8 for an image of the two extreme conditions). To ensure that the virtual peer learners' hand-raising behavior was unambiguously attributed to their performance level (i.e., more hand-raising peers leading to the perception of a higher-performing class), the hand-raising virtual classmates' answers were always correct, which was communicated accordingly by the virtual teacher. A manipulation check indicated that virtual classmates' hand-raising was significantly positively related to the perceived performance level of the class (assessed via self-reports from participants after the IVR lesson; Spearman's rho $\rho = .41$, $p < .001$). Mean differences in the perceived performance level continuously increased from 20% hand-raising ($M = 2.89$, $SD = 0.53$) to 80% hand-raising ($M = 3.45$, $SD = 0.43$).

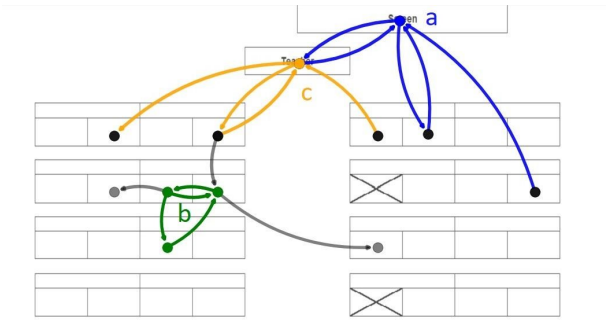
Overall IVR classroom design and virtual avatar animation

All configuration conditions aimed at a high degree of behavioral realism [425], [426] and considered the Uncanny Valley effect in terms of appropriate avatar visualization [400], [402]. To ensure that the IVR lesson was consistent with a typical classroom experience for sixth graders and perceived as authentic, we used audio recordings and motion captures from a real sixth-grade classroom. We recorded and motion-captured six different students for the duration of the whole 15-minute IVR lesson duration in a regular school setting. That is, we equipped six students with a motion-capturing suit during a regular lesson at their school; to record authentic movements, we asked these students to behave like they usually would in their classroom. We then used individual sequences of the recorded movements to distinctly animate the simulated virtual peer learners in the classroom. These individual sequences consisted of, for instance, different postures while sitting at the table (e.g., leaned back with hands in the lap or on the table or very upright with arms propped up on the table), different body movements (e.g., swinging their feet or shifting their weight back and forth on the chair), different head and shoulder movements (e.g., quickly scanning the classroom or slowly turning the head back and forth), or different hand-raising styles (e.g., lifting one arm straight upward or supporting the hand-raising arm with the other hand propped up on the

table).

Notably, participants reported similarly high levels of perceived realism and experienced presence in the IVR classroom across all configuration conditions. We assessed participants' perceived realism and experienced presence in the IVR classroom with six and nine items each in the posttest questionnaire (see Appendix B for the full set of items). The measures were based on conceptualizations of presence by Schubert et al. [79] and Lombard et al. [317] and adapted to assess students' perception of and experience with the specific IVR environment in the present study. Both perceived realism (e.g., "What I experienced in the virtual classroom could also happen in a real classroom") and experienced presence (e.g., "I felt like I was sitting in the virtual classroom") were rated on 4-point rating scales ranging from 1 to 4, with higher values indicating higher levels of perceived realism and experienced presence (Cronbach's alpha values of 0.76 and 0.77, respectively). The IVR configuration had no statistically significant effect on participants' perceived realism and experienced presence; mean values ranged between 2.73 and 3.08 ($0.26 < SDs < 0.75$) for both variables across all IVR conditions.

Figure C.9.: Example Visualization of Structural Graph Variables



Note. Three different structural variables are visualized for illustrative purposes, one from each respective category. a = gaze centrality of the screen (visualized in blue): all incoming and outgoing edge weights are summed up into one centrality marker. b = clique among peers (visualized in green): all green nodes are connected to each other by edges and therefore form a clique; a connection is considered to exist if there is at least one edge between two nodes. c = cut size between teacher and peers (visualized in yellow): summing up all edge weights between the teacher and all peer nodes.

IVR lesson content and procedure

The IVR experience was a 15-min simulation of a teacher-directed lesson on computational thinking designed for sixth-grade students. We chose computational thinking as the IVR lesson content because—despite being considered a central 21st century skill—this topic is not yet widely taught in primary and early secondary education, except for in some extracurricular activities [316], [427]. Aiming to provide participants with novel content they had little (or no) prior knowledge or learning experiences with, the 15-min IVR lesson's goal was to introduce sequences and loops as basic computational concepts. Participants experienced the IVR lesson from the perspective of a student in the IVR classroom surrounded by 24 simulated virtual peer learners. The IVR lesson proceeded in the fashion of a typical classroom situation consisting of two phases directed by the teacher. In the first phase, the virtual teacher introduced the topic and central concepts and asked questions to include the students (e.g., "Can anyone explain to me what programming means and is good for?" or "Do you know an example from your daily life of something that works like a sequence or loop?"). The virtual peer learners were programmed to raise their hands in response to the teacher's questions and to respond when the virtual teacher called on them. Participants could raise their hands; however, they were not called on by the teacher since the whole IVR lesson was fully preprogrammed. In the second phase, the virtual teacher presented two exercises that students had some time to think about individually, and lastly, the virtual teacher discussed the solutions to these exercises. A detailed schedule of the IVR lesson is provided in the supplemental material. The participating students had no individual learning materials in the IVR, but the virtual teacher referred to slides on the screen in the front of the classroom on which central definitions (Phase 1) and the exercises (Phase 2) were presented. The instructional goals, content, and approach were exactly the same in all IVR configurations; the experimental conditions differed only with respect to the position from which participants experienced the lesson, the visualization style of the virtual teacher and peer learners, and the proportion of hand-raising peers when a question was asked (see Figures C.7 and C.8).

Measures

Structural variables describing gaze-based attention networks

In order to generate gaze-based attention networks from the gaze and head movement data, various data pre-processing steps were necessary. A detailed description of the preprocessing of eye-tracking data and the creation of graphs is provided with the analysis

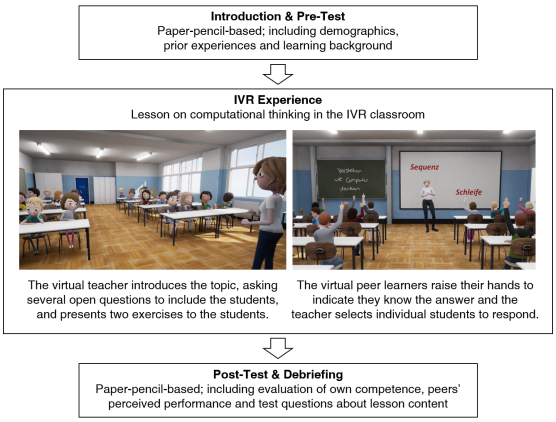
procedure in Methods Section 3.6.1. First, we used a technique known as gaze ray-casting [179], [195] to extract the information about what participants looked at during the IVR lesson. Because not every virtual object in the environment was of interest for our study, we afterward defined and only included specific objects of interest (OOIs) with regard to our research questions, namely the virtual peer learners, the virtual teacher, and the screen on which the instructional content was presented. We then counted participants' gaze transitions between these OOIs as gaze shifts and aggregated the number of gaze shifts across all OOIs during the whole experiment. The collected gaze information was then transformed into a separate graph that treated (a) the observed OOIs (i.e., the teacher, screen, and peer learners that were gazed at) as nodes and (b) gaze shifts between the OOIs as edges. The graph was constructed with bidirectional connections between these nodes as weighted edges (see Figure 3 for a visualization). Each weighted edge was defined by the number of gaze shifts from one OOI to another during the whole experiment (i.e., how often participants' gaze shifted between the two nodes that the edge connected). We refer to this frequency of gaze shifts as edge weight.

Based on these graphs for each participant for the full experiment duration, we calculated different structural variables that describe the graph structure and associated gaze-based attention network for each participant. The calculated structural variables can be assigned to three categories, namely gaze centrality, connectedness of gazes and uniformity of gaze distribution.

Gaze centrality was assessed via three variables regarding central OOIs in the IVR classroom: the degree centrality of (a) the peer learners, (b) the virtual teacher, and (c) the screen with the instructional content. Degree centrality [145], [146], as a measure of the gaze centrality of the OOIs, indicates to what extent these OOIs are in the center of gaze transitions and describes the focus of attention towards these OOIs. For each node (or bundle of nodes) in the graph, degree centrality is defined as the sum of weights of all incoming and outgoing edges (or the sum of all edge weights for more than one node). To calculate degree centrality in our gaze-based attention networks, we summed up the frequency of gaze shifts from and towards the selected OOIs.

Connectedness of gazes was measured with three variables regarding so-called cliques in the gaze network. Cliques are highly connected clusters (i.e., substructures) in a graph and therefore provide information about the connectedness of gazes—i.e., the extent of gaze transitions between the OOIs—in the gaze-based attention network. Because cliques can only be calculated in undirected graphs, we transformed each directed graph into an

Figure C.10.: Study Procedure and IVR Lesson Content



Note. The images depict a situation during the IVR lesson when the virtual teacher asked a question and virtual students raised their hands to indicate that they know the answer. The image on the left shows virtual avatars represented in a cartoon-style manner (and with less hand-raising), the image on the right depicts stylized virtual avatars (with more hand-raising). Participants experienced the classroom situation either from the second row from the front or the back row.

undirected one by calculating the weight of each undirected edge as the sum of both directed edge weights. Furthermore, we calculated all maximal cliques among virtual peer learners (i.e., the subset of nodes that contains the maximal number of nodes that share an edge with every other node in the subset). Because two connected nodes build a trivial clique, we only considered cliques that contain more than two nodes. After calculating all cliques in a network, we are able to state (a) the number of cliques, and (b) their average size. Given that this part of the analysis focused specifically on visual attention towards virtual peer learners, we took this opportunity to conduct more fine-grained analyses, such as the gender composition of cliques. Therefore, we calculated (c) the proportion of boys in the observed cliques.

Uniformity of gaze distribution was measured with three different variables that describe how gaze shifts were distributed across the OOIs in the classroom. Firstly, we calculated (a) a weighted degree centrality measure (as proposed by [183]) that includes uniformity of edge weights. We consequently used the weighted degree centrality (WDC) of the screen as an indicator of how uniformly gaze was distributed from the screen to different peer learners. Secondly, we calculated (b) the cut size between the teacher/screen and peer learners as an

indicator of how much students' visual attention shifted back and forth between the two versus staying on one group (e.g., students mostly focused on the teacher/screen). Cut size was calculated by summing up the edge weights of edges that pass between the two subsets (one subset being the teacher and screen and the other being all peer learners). Thirdly, we looked at the overall distribution of all edge weights in the network and tested for (c) overall uniformity. Therefore, for each participant, we stated the chi-square test statistic value calculated for a sample containing all edge weights of this person's gaze shifts.

Students' learning outcomes

Interest. Participants' interest in the IVR lesson was measured at posttest with six items (e.g., "I liked the topic of the lesson" or "I would like to learn more about the topic of the lesson"; based on [428]) on a 4-point rating scale ranging from 1 (*not true at all*) to 4 (*absolutely true*), yielding a Cronbach's alpha of .91 ($M = 3.18$, $SD = 0.69$).

Situational self-concept. Participants' situational self-concept after the IVR lesson was assessed with a four-item scale that was based on the commonly used wording by Schwanzer, Trautwein, Lüdtke, and Sydow [429] and adapted for the specific situation with virtual peer learners (e.g., "I could solve the robot tasks faster than the others" and "It was harder for me to understand the robot tasks than for the other students"). Participants indicated their responses on a 4-point rating scale ranging from 1 (*not true at all*) to 4 (*absolutely true*). Two items were reverse-scored and recoded accordingly, yielding an acceptable Cronbach's alpha of .69 in our sample overall ($M = 3.41$, $SD = 0.54$).

Learning. The posttest questionnaire included a short test of how much participants learned during the IVR lesson about computational thinking (based on the Computational Thinking test by [430]). The test consisted of 12 questions covering the IVR lesson content on basic computational principles (i.e., sequences and loops). Participants had to indicate whether 12 given statements were correct or incorrect (e.g., "The order of commands does not matter in a loop" [false] or "Following a recipe when cooking is an example of a sequence" [correct]). Participants were given one point for each correct answer; thus, posttest scores ranged from 0 to a maximum of 12 points. Obtained scores ranged from 4 to 12 ($M = 10.45$, $SD = 1.59$). The 12 items had a low but acceptable Cronbach's alpha of .53.

Appendix B shows the complete instruments and all items used to assess students' learning outcomes.

Covariates

Participants' gender, general intelligence self-concept (based on [429]) and initial interest in the lesson topic of computational thinking were included as covariates in the models, as they could potentially influence participants' learning outcomes. Intelligence self-concept was measured with four items (e.g., "I often think I'm not as smart as the others"), and initial interest in the lesson topic was measured with five items (e.g., "I would like to know more about how computer applications and robots work" or "I am interested in topics related to technology"). The full instruments are included in Appendix B. A four-point rating scale ranging from 1 (*not true at all*) to 4 (*absolutely true*) were used for both scales, yielding Cronbach's alpha values of .74 for intelligence self-concept ($M = 3.07$, $SD = 0.61$) and .91 for initial interest in the lesson topic ($M = 3.14$, $SD = 0.76$).

Data collection

Apparatus

We used HTC Vive Pro Eye head-mounted displays with a refresh rate of 90 Hz and a 110° field of view for our experiment (1440 x 1600 screen resolution for each eye). To collect participants' eye-tracking data during the IVR lesson, we used the integrated Tobii eye tracker in the HTC Vive Pro Eye with a 120 Hz sampling rate and a default calibration accuracy of 0.5° – 1.1° (based on a standard 5-point calibration). The IVR classroom scenario was designed and rendered using the Unreal Game Engine v4.23.1.

Study procedure

Participants took part in the experiment in a quiet room at their school in groups of up to 10. Importantly, whereas up to 10 participants experienced the IVR classroom situation at the same time, their IVR systems were not linked in any way. That is, all virtual characters in the classroom (i.e., the teacher and all peer learners) were simulated and fully preprogrammed avatars, and each participant experienced the IVR classroom situation surrounded by the same 24 simulated virtual peer learners. Each of the test sessions followed the same procedure, which consisted of three parts (see Figure 4 for an overview).

After a general introduction to the study procedure, participants completed a paper-based pretest questionnaire including demographics and relevant background variables (e.g., intelligence self-concept, initial interest in the lesson topic and previous IVR experience). Following the pretest, participants put on the head-mounted displays and experienced the 15-minute IVR lesson once the integrated eye tracker was calibrated. The IVR lesson was introduced

as a learning experience that participants were free to explore as they liked. Participants were seated in desks in the real world, congruent to their virtual IVR classroom experience; they were instructed to remain seated but otherwise behave like they would in a normal classroom situation. Participating students were unaware of the different IVR configuration conditions during the experiment. Upon completion of the IVR lesson, participants filled out the paper-based posttest questionnaire, including measures of their self-concept with reference to the IVR situation and their overall experience of the IVR experience. The testing session ended with a debriefing about the study aims and design (including information about the random assignment to different IVR configurations) after approximately 45 min in total.

Analysis procedure

Pre-processing of eye-tracking data and creation of graphs

We analyzed attentional processes by performing network analysis on gaze shift movements to trace the path of visual attention throughout the virtual space. To access information about where participants looked during the IVR lesson, we used a technique known as gaze ray-casting [179], [195]. Gaze ray-casting combines information from each frame about the participant's head location, head orientation, and gaze direction to calculate the gaze direction in the virtual environment and pinpoint the exact location the participant is looking at. One could imagine a gaze ray-cast as a laser beam pointing from the participant's (combined) eye location into the virtual space and hitting a specific physical object there. By identifying hits of the gaze ray with a virtual object at every split second, we were able to continuously track which object in the IVR classroom participants observed [179]. We used the Python programming language (Python Software Foundation, `https://www.python.org/`) to process the ray-casting information calculated during the experimental session. Since not every virtual object in the environment was of interest for our study, we only included specific objects of interest (OOIs), namely the virtual peer learners, the virtual teacher, and the screen on which the instructional content was presented. The gaze shift movement from one object to another was then identified as a transition from a specific object to another if the transition duration was no longer than 10 seconds. The transitions between OOIs across the entire experiment session were then summed up and stored in an adjacency matrix (i.e., a $n \times n$ matrix A, with n being the number of OOIs). Consequently, each cell in the matrix a_{ij} stated the number of gaze shifts transitioning from OOI i to OOI j, resulting in a transition matrix for each participant encompassing the number of gaze shifts for the full experiment session

(similar to transitions matrices for scanpath analyses; [114]).

According to graph theory, each adjacency matrix can be considered as a weighted directed cyclic graph [340]. Building a graph from the normalized transition matrix involves treating the OOIs as nodes of the graph and the normalized number of transitions as the edge weights, with an edge created between nodes if at least two transitions between them occurred during the full experiment. We built the graphs from the adjacency matrices using the NetworkX package [351]. Based on a graph for each participant, we were able to calculate different variables describing the graph structure and graph properties (see overview in Appendix A and in Measures Section 3.4.1).

Notably, the gaze-based graphs also include aggregated information about head movements; because participants' field of view did not capture all OOIs in the classroom simultaneously, even when participations were positioned in the back of the IVR classroom, the gaze-based graphs include information about the extent to which participants moved their heads to the left and right to change their field of view (e.g., high gaze activity among peers is only possible with head movements).

We provide access to all data and analysis scripts for the data pre-processing steps on the Open Science Framework (OSF) under the following link: `https://osf.io/pek4q/?view_only=ef151fd06ac8413a827020d4264b3c8d`.

Statistical analyses

We applied three-way full factorial ANOVAs to examine differences in structural variables of students' gaze-based attention networks in the different IVR configuration conditions (RQ1, H1a-d). To answer RQ2, we used partial correlations to examine the relation between structural variables of students' gaze-based attention networks and (a) their interest in the IVR lesson, (b) their situational self-concept, and (c) their posttest learning score after the IVR lesson. We added students' gender, general intelligence self-concept and initial interest in the lesson topic as covariates, to account for the fact that these variables could potentially influence students' learning outcomes (a respective correlation matrix of the covariates and outcome variables is provided in the supplemental material). Moreover, as we sought to obtain insights into the general meaning of the gaze features used, we controlled for the IVR configuration conditions (i.e., participants' position in the classroom, the virtual avatars' visualization styles, and virtual peers' performance-related behavior) to examine the relations between students' visual attention and their learning experiences after removing the influence of our experimental manipulation.

Prior to all analyses, we checked for a normal distribution of our data with the Shapiro-Wilk Test. If the Shapiro-Wilk test was significant, and graphical representations and variable skewness and kurtosis also indicated a lack of normality, we calculated Spearman's rho for non-parametric correlations. For non-parametric ANOVA procedures, we applied full-factorial aligned rank transformation using the ARTool package in R [319]. We used Tukey's HSD test for post-hoc comparisons and calculated partial eta squared (η_p^2) to describe effect sizes of the ANOVA, with cut-off values of ≥ 0.06 for medium and ≥ 0.14 for large effects; in addition, we report Cohen's d for all main results, with cut-off values of ≥ 0.5 for medium and ≥ 0.8 for large effects [431]. All analyses were done in R [432] and we set the critical p-value and confidence intervals at an alpha level of .05 for all hypothesis tests. We report and interpret results based on both statistical significance and effect sizes.

We posted all data and data analysis scripts on the Open Science Framework under the following link: `https://osf.io/pek4q/?view_only=ef151fd06ac8413a827020d4264b3c8d`.

C.2.5. Results

We used different structural features to analyze students' gaze-based attention networks describing gaze centrality, the connectedness of gazes among peers and the overall uniformity of gaze distribution. Figure C.10 depicts examples of gaze-based attention networks from selected students. Table C.5 provides descriptive statistics and a correlation matrix for the structural features describing students' gaze-based attention networks. We found significant moderate to high correlations between almost all the features. On the one hand, high correlations are to be expected for markers that are based on similar calculations, such as degree centrality of peers, teacher and screen, or the number and average size of cliques—particularly in light of the highly standardized environment. On the other hand, each of the selected features provides distinct information about the visual attention toward (social) information in the classroom (see Figure C.9). As also depicted in Figure C.11, the correlational pattern between the structural variables indicated that students' gaze-based attention networks largely reflect two types: Students tended to focus their gazes either on their peers or on the instructional content (i.e., the teacher and screen; see highly negative correlation of degree centrality peers with degree centrality teacher, $\rho = -0.93$, $p < .001$, and degree centrality screen, $\rho = -0.91$, $p < .001$). The more students' gaze centered on the instructional content (i.e., the teacher or screen), the less uniformly their gazes were distributed across the classroom, as can be seen, for instance, in the comparably high negative correlations

of the uniformity of gaze distribution between the screen and peers (i.e., WDC screen) with the degree centrality of the teacher ($\rho = -0.87$, $p < .001$) and screen ($\rho = -0.64$, $p < .001$), in contrast to highly positive correlations with the degree centrality of peers ($\rho = 0.80$, $p < .001$).

Table C.5.: Descriptive Statistics and Correlation Matrix for Structural Variables Describing Students' Gaze-Based Attention Networks.

Variable	*M (SD)*	*Mdn (MAD)*	*Min / Max*	1	2	3	4	5	6	7	8
1. DC peers	0.72 (0.35)	0.73 (0.39)	0.04 / 1.74	—							
2. DC teacher	0.58 (0.20)	0.58 (0.21)	0.02 / 0.97	-.93*** [-0.95, -0.91]	—						
3. DC screen	0.71 (0.18)	0.73 (0.18)	0.12 / 1.00	-.91*** [-0.93, -0.88]	.73*** [0.66, 0.79]	—					
4. N cliques peers	2.78 (2.75)	2.00 (2.97)	0.00 / 12.00	.59*** [0.50, 0.67]	-.48*** [-0.57, -0.38]	-.67*** [-0.74, -0.59]	—				
5. Avg clique size peers	2.31 (0.71)	2.38 (0.56)	0.00 / 3.58	.56*** [0.47, 0.64]	-.48*** [-0.57, -0.38]	-.62*** [-0.69, -0.53]	.92*** [0.90, 0.94]	—			
6. Proportion boys in cliques	0.53 (0.19)	0.56 (0.13)	0.00 / 1.99	.57*** [0.48, 0.65]	-.53*** [-0.62, -0.43]	-.54*** [-0.62, -0.44]	.35*** [0.24, 0.45]	.34*** [0.23, 0.44]	—		
7. WDC screen	1.93 (0.54)	1.89 (0.68)	1.05 / 4.00	.80*** [0.75, 0.84]	-.87*** [-0.90, -0.83]	-.64*** [-0.71, -0.56]	.40*** [0.29, 0.50]	.40*** [0.29, 0.50]	.53*** [0.43, 0.62]	—	
8. CS teacher/ screen – peers	1.78 (0.57)	1.75 (0.63)	0.57 / 3.45	.51*** [0.41, 0.60]	-.60*** [-0.68, -0.51]	-.28*** [-0.39, -0.16]	-.07 [-0.19, 0.05]	-.06 [-0.18, 0.06]	.23*** [0.11, 0.34]	.60*** [0.51, 0.68]	—
9. Uniformity overall GD	-4.53 (2.33)	-4.21 (2.43)	-11.27 / -0.35	.36*** [0.25, 0.47]	-.36*** [-0.46, -0.25]	-.25*** [-0.35, -0.13]	-.25*** [-0.36, -0.13]	-.24*** [-0.35, -0.12]	.11 [-0.22, 0.01]	.30*** [0.19, 0.41]	.67*** [0.59, 0.74]

Note. Mean values and standard deviations (M and SD) as well as medians and median absolute deviations (Mdn and MAD) are reported. Variables 1-9 are non-normally distributed; thus, Spearman's rho is reported. 95% confidence intervals are given in brackets. DC = Degree Centrality; N = Number; Avg = Average; WDC = Weighted Degree Centrality; CS = Cut Size; GD = Gaze Distribution. *** $p < .001$.

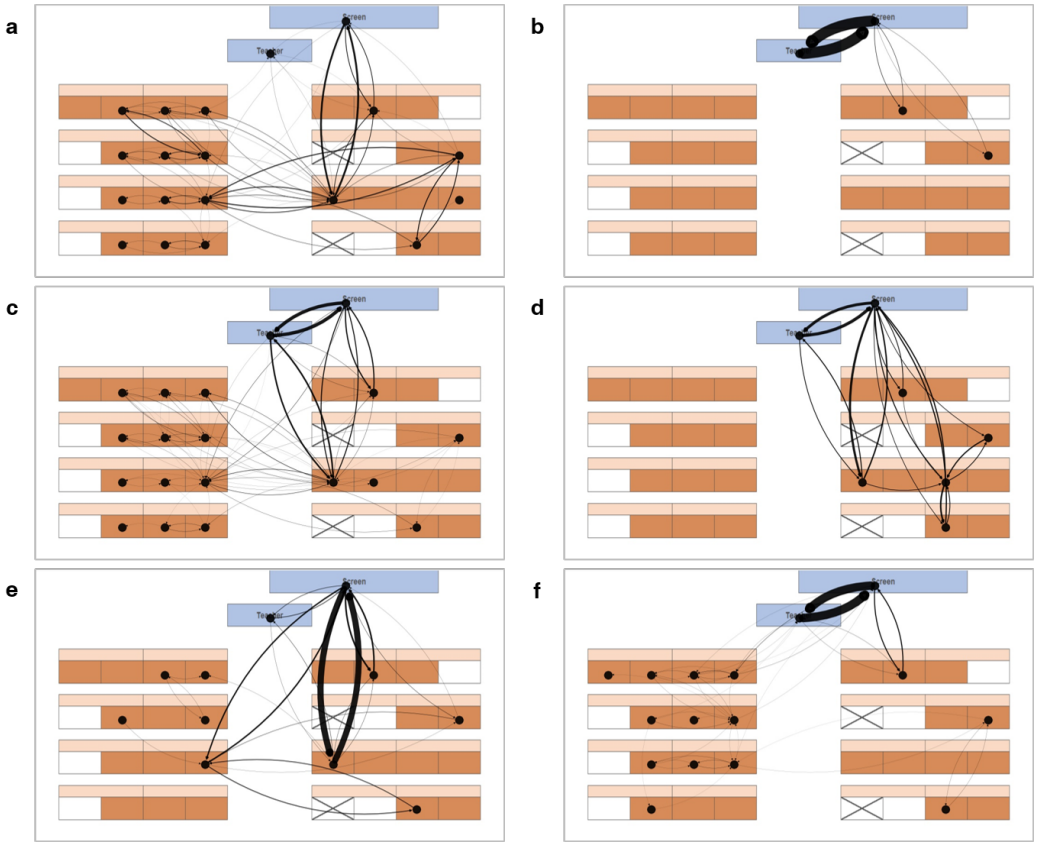
How do different social-related IVR configurations affect the structure of students' gaze-based attention networks in the IVR classroom?

The first research question (RQ1) asked how the social-related IVR classroom configurations (i.e., participants' position, visualization style of virtual avatars, and virtual peers' performance-related behavior) affect students' gaze-based attention networks during the IVR lesson. In this section, we describe the results of three-way full-factorial aligned rank transformation ANOVAs examining the effects of the IVR configurations on the structural features of students' gaze-based attention networks (i.e., gaze centrality, connectedness of gazes among peers, uniformity of gaze distribution). We only report detailed statistics for significant results in the main text; full statistics for all analyses are provided in the supplemental material. Table C.6 provides an overview of the observed main effects.

Effects of students' seating position on the structure of their gaze-based attention networks

Our first hypothesis (H1a) was that students who were positioned in the back of the IVR classroom would show more gaze centrality on virtual peer learners (and less centered gaze networks on the virtual teacher or screen), more connectedness of gazes among peers, and

Figure C.11.: Example Gaze-Based Attention Networks for Different Participants



Note. a–f represent the gaze-based attention networks over the course of the 15-minute IVR lesson for six selected participants. The crossed-out seats indicate the participants' position in the front (b, c, and f) versus the back (a, d and e) of the classroom. Colored seats were occupied by a virtual peer learner, white seats were empty. Black bullets represent nodes (i.e., OOIs gazed at); the width of the black lines indicates the frequency of gaze transitions between the OOIs. a = high gaze centrality of peers, low gaze centrality of teacher and screen. b = low gaze centrality of peers (no cliques), high gaze centrality of teacher and screen, low weighted degree centrality screen (uniformity in gaze distribution between screen and peers) and low cut size (transitions between teacher/screen and peers). c = high number and average size of cliques among peers. d = high weighted degree centrality screen and high uniformity of gaze distribution across all OOIs, as indicated by similar width of all connecting lines. e = High cut size and medium uniformity of gaze distribution across OOIs. f = Low uniformity and cut size, medium number and size of cliques.

Table C.6.: Summary of Main Effects of IVR Configuration Conditions on Structural Network Features

	Seating position	Avatar visualization style	Proportion of hand-raising peer learners
DC peers	**front <back**	**cartoonish >stylized**	20% >65% <80%
DC teacher	**front >back**	**cartoonish <stylized**	20% <65% >80%
DC screen	**front >back**	**no difference**	20% <65% >80%
N cliques peers	**front <back**	**cartoonish >stylized**	no difference
Avg clique size peers	**front <back**	**cartoonish >stylized**	no difference
Proportion boys in cliques	**front <back**	no difference	no difference
WDC screen	**front <back**	**cartoonish >stylized**	20% >35% >65% <80%
CS teacher/screen – peers	**front <back**	**cartoonish >stylized**	no difference
Uniformity overall	**front <back**	no difference	no difference

Note. Only statistically significant differences are shown. < and > indicate the direction of the effect. Results in bold represent findings in line with hypotheses. DC = Degree Centrality; N = Number; Avg = Average; WDC = Weighted Degree Centrality; CS = Cut Size; GD = Gaze Distribution.

due to the increased field of view, a more uniformly distributed gaze in the IVR classroom. As expected, students' position in the IVR classroom had a significant effect on all of the structural features describing students' gaze-based attention networks (descriptive statistics in Table C.7).

First, students in the front position showed a significantly different focus of gaze transitions on their virtual peer learners, the virtual teacher and the screen with the lesson content (i.e., measured by degree centrality) compared to students in the back seating position: Degree centrality of the virtual peers was significantly higher when students were located in the back row of the IVR classroom, $F(1,258) = 138.55$, $p < .001$, $\eta_p^2 = 0.35$, $d = 1.36$. In turn, degree centrality of the virtual teacher was significantly higher for students who were located in the front of the IVR classroom, $F(1,258) = 204.07$, $p < .001$, $\eta_p^2 = 0.44$, $d = 1.53$, and similarly the screen was more the focus of students' gaze transitions when they were sitting in the front, $F(1,258) = 60.56$, $p < .001$, $\eta_p^2 = 0.19$, $d = 0.90$.

Second, with regard to the gaze activity among virtual peer learners, we found that the front vs. back position in the IVR classroom led to significant differences in the number and average size of cliques and the proportion of boys in the observed cliques: The number of cliques among peers and the average size of these cliques were significantly higher when students were positioned in the back of the IVR classroom, whereby also the proportion of boys in the observed cliques was significantly higher in the back position; $F(1,258) = 4.50$, $p = .035$, $\eta_p^2 = 0.02$, $d = 0.38$ and $F(1,258) = 8.06$, $p = .005$, $\eta_p^2 = 0.03$, $d = 0.28$, and $F(1,258) = 101.35$, $p < .001$, $\eta_p^2 = 0.28$, $d = 0.85$ for the number and average size of cliques and the proportion of boys in the observed cliques, respectively.

Third, results showed higher levels of uniformity in gaze distribution for students in the back seating position for all three indicators of uniformity: Students distributed their gazes more evenly from the screen to different peers (i.e., weighted degree centrality screen) in the back position, $F(1,258) = 333.16$, $p < .001$, $\eta_p^2 = 0.56$, $d = 2.04$, and had more gaze transitions between the instructional content (teacher/screen) and peers (i.e., cut size) when sitting in the back, $F(1,258) = 109.23$, $p < .001$, $\eta_p^2 = 0.30$, $d = 1.22$. Moreover, students' gazes were distributed more uniformly across all observed OOIs in the IVR classroom when they were sitting in the back; $F(1,258) = 33.11$, $p < .001$, $\eta_p^2 = 0.11$, $d = 0.65$.

Table C.7.: Descriptive Statistics for Structural Network Features in Different Seating Positions

	Front (N = 122)		Back (N = 152)	
	M (SD)	*Mdn (MAD)*	*M (SD)*	*Mdn (MAD)*
DC peers	0.50 (0.27)	0.45 (0.31)	0.89 (0.30)	0.89 (0.25)
DC teacher	0.71 (0.14)	0.72 (0.16)	0.47 (0.17)	0.46 (0.17)
DC screen	0.79 (0.15)	0.83 (0.14)	0.64 (0.18)	0.65 (0.15)
N cliques peers	2.21 (2.39)	2.00 (2.97)	3.24 (2.93)	3.00 (4.45)
Avg clique size peers	2.20 (0.70)	2.32 (0.48)	2.40 (0.71)	2.44 (0.58)
Proportion boys in cliques	0.45 (0.17)	0.50 (0.11)	0.60 (0.18)	0.63 (0.09)
WDC screen	1.50 (0.28)	1.43 (0.22)	2.27 (0.44)	2.29 (0.39)
CS teacher/screen – peers	1.45 (0.47)	1.38 (0.43)	2.05 (0.51)	2.06 (0.50)
Uniformity overall	-5.34 (2.52)	-5.28 (2.72)	-3.89 (1.94)	-3.86 (1.96)

Note. Mean values and standard deviations (M and SD, respectively) as well as medians and median absolute deviations (Mdn and MAD, respectively) are reported. Values for each of the configuration conditions are averaged across the other conditions. DC = Degree Centrality; N = Number; Avg = Average; WDC = Weighted Degree Centrality; CS = Cut Size; GD = Gaze Distribution.

In sum, these results fully support our Hypothesis H1a: The position in the back of the virtual classroom led to more gaze centrality on virtual peer learners (and less on the virtual teacher and screen), more connectedness of gazes among peers and more uniformly distributed gazes between OOIs in the IVR classroom (see summary in Table C.6).

Effects of virtual avatar visualization style on the structure of students' gaze-based attention networks

Our second hypothesis (H1b) suggested that a cartoonish visualization of avatars would lead to more gaze centrality on virtual peer learners and less gaze centrality on the virtual teacher (gaze centrality on the screen not affected), more connectedness of gazes among peers, and due to the increased focus on peer learners, a less uniformly distributed gaze

overall in the IVR classroom. Results showed that the virtual avatars' visualization style had a significant effect on a number of the structural features describing students' gaze-based attention networks (descriptive statistics in Table C.8).

Table C.8.: Descriptive Statistics for Structural Network Features in Different Avatar Visualization Styles

	Cartoonish (N = 150)		Stylized (N = 124)	
	M (SD)	*Mdn (MAD)*	*M (SD)*	*Mdn (MAD)*
DC peers	0.79 (0.35)	0.80 (0.36)	0.63 (0.32)	0.60 (0.38)
DC teacher	0.52 (0.21)	0.49 (0.22)	0.65 (0.16)	0.63 (0.17)
DC screen	0.69 (0.18)	0.71 (0.17)	0.73 (0.19)	0.76 (0.19)
N cliques peers	3.24 (2.80)	3.00 (2.97)	2.23 (2.59)	1.00 (1.48)
Avg clique size peers	2.44 (0.62)	2.48 (0.50)	2.15 (0.79)	2.26 (0.39)
Proportion boys in cliques	0.54 (0.16)	0.57 (0.12)	0.52 (0.22)	0.55 (0.13)
WDC screen	2.04 (0.54)	2.08 (0.69)	1.80 (0.51)	1.70 (0.52)
CS teacher/screen – peers	1.88 (0.60)	1.87 (0.65)	1.66 (0.50)	1.66 (0.50)
Uniformity overall GD	-4.56 (2.43)	-4.17 (2.33)	-4.50 (2.20)	-4.32 (2.54)

Note. Mean values and standard deviations (M and SD, respectively) as well as medians and median absolute deviations (Mdn and MAD, respectively) are reported. Values are averaged across the other configuration conditions. DC = Degree Centrality; N = Number; Avg = Average; WDC = Weighted Degree Centrality; CS = Cut Size; GD = Gaze Distribution.

With regards to the degree to which OOIs were at the focus of students' gaze transitions (i.e., measured by degree centrality), we found greater degree centrality of peer learners when they were presented in cartoon style compared to a more stylized visualization, $F(1,258) = 20.46$, $p < .001$, $\eta_p^2 = 0.07$, $d = 0.48$. In contrast, degree centrality of the teacher was higher in the stylized visualization compared to the cartoonish one, $F(1,258) = 52.03$, $p < .001$, $\eta_p^2 = 0.17$, $d = 0.69$. We found no statistically significant differences for the degree centrality of the screen based on different avatar visualizations.

Turning to the gaze activity among virtual peer learners, the number and the average size of cliques among peers differed significantly between the visualization styles of the virtual avatars in the IVR. Both the number of cliques among peers and the average clique size were statistically significantly higher when virtual peer learners were visualized in cartoon style; $F(1,258) = 9.81$, $p = .002$, $\eta_p^2 = 0.04$, $d = 0.37$, and $F(1,258) = 12.47$, $p < .001$, $\eta_p^2 = 0.05$, $d = 0.41$ for the number and average size of cliques, respectively. The proportion of boys in the observed cliques was not affected by the visualization style of virtual avatars.

Moreover, we found more evenly distributed gazes between the screen and peers (i.e.,

higher weighted degree centrality screen) when avatars were visualized in cartoon style, $F(1,258) = 34.82$, $p < .001$, $\eta_p^2 = 0.12$, $d = 0.46$. Similarly, the results showed more gaze transitions between the teacher/screen and peers (i.e., higher cut size) for the cartoonish visualization, $F(1,258) = 14.30$, $p < .001$, $\eta_p^2 = 0.05$, $d = 0.39$. The visualization style of virtual avatars had no effect on the overall uniformity of gaze distribution across OOIs in the IVR classroom.

In sum, these results partially support our Hypothesis H1b (see summary in Table C.6).

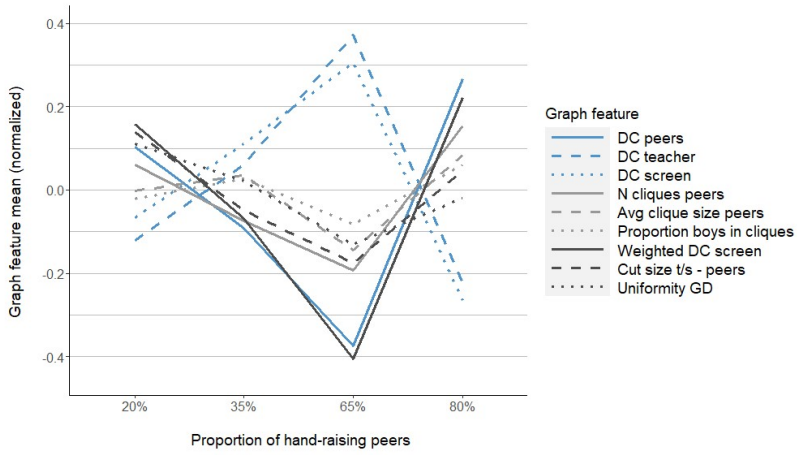
Effects of virtual peers' hand-raising behavior on the structure of students' gaze-based attention networks

Our third hypothesis (H1c) was that more hand-raising behavior of virtual classmates would lead to more gaze centrality on peer learners (and gaze networks less centered on the teacher or screen), more connected gazes among peers, and based on the desire to obtain a comprehensive picture of peer learners' behavior, a more uniformly distributed gaze in the IVR classroom. In fact, the manipulation of virtual peers' performance-related behavior, specifically their hand-raising behavior, had a statistically significant effect on students' gaze-based attention networks. In particular, the degree to which virtual peer learners, the virtual teacher and the screen with the lesson content were the focus of students' gaze transitions (measured by degree centrality) was affected by the hand-raising conditions; $F(3,258) = 7.76$, $p < .001$, $\eta_p^2 = 0.08$ and $F(3,258) = 8.09$, $p < .001$, $\eta_p^2 = 0.09$ and $F(3,258) = 4.94$, $p = .002$, $\eta_p^2 = 0.05$ for the degree centrality of virtual peer learners, the virtual teacher and the screen, respectively. As can be seen in Figure C.12, descriptively speaking, the degree centrality of virtual peers was highest in the 'extreme' hand-raising conditions of 20% and 80% (see solid blue line), whereas the degree centrality of the teacher and the screen showed the opposite pattern (see dotted and dashed blue lines).

Tukey's HSD test for multiple comparisons showed that the degree centrality of peers was significantly higher in the 20% hand-raising condition compared to the 65% condition ($p = .004$, $d = 0.48$) and significantly lower for the 65% compared to the 80% hand-raising condition ($p < .001$, $d = 0.64$). In turn, the degree centrality of the teacher and screen were significantly lower in the 20% hand-raising condition compared to the 65% condition ($p = .002$, $d = 0.54$, and $p = .039$, $d = 0.34$ for the teacher and screen, respectively) and significantly higher for 65% compared to 80% hand-raising ($p < .001$, $d = 0.63$, and $p = .002$, $d = 0.55$ for the teacher and screen, respectively).

With regard to gaze activity among virtual peer learners, the hand-raising conditions had

Figure C.12.: Normalized Mean Values of Gaze Features by Hand-Raising Conditions



Note. Values are averaged across seating position and avatar visualization. DC = Degree Centrality; N = Number; Avg = Average; t/s = teacher/screen; GD = Gaze Distribution.

no statistically significant effect on any of the respective features (i.e., number and average size of cliques as well as proportion of boys in the observed cliques; see light grey lines in Figure C.12).

Lastly, whereas the hand-raising behavior of virtual peers also had no effect on the amount of transitions between teacher/screen and peers (i.e., cut size) and the overall uniformity of gaze distribution across OOIs in the IVR classroom, the results indicated significantly different levels of weighted degree centrality of the screen in the different hand-raising conditions, $F(3,258) = 12.92$, $p < .001$, $\eta_p^2 = 0.13$. A similar pattern as for the gaze centrality markers was observed: Weighted degree centrality of the screen (i.e., uniformity of gaze distribution between screen and peers) was highest in the 20% and 80% hand-raising conditions (see solid dark grey line in Figure 6). Tukey's HSD test for multiple comparisons indicated statistically significant differences between 20% and 65% ($p < .001$, $d = 0.60$), 35% and 65% ($p = .028$, $d = 0.35$), 35% and 80% ($p = .017$, $d = 0.29$), and 65% and 80% ($p < .001$, $d = 0.69$).

In sum, the pattern of results for the effects of peer learners' hand-raising on students' gaze-based attention networks in the IVR classroom was different than we hypothesized (H1c). Descriptive statistics for the structural features of students' gaze-based attention networks in the different hand-raising conditions are given in Table C.9. Detailed statistics

for the post-hoc comparisons can be found in the supplemental material.

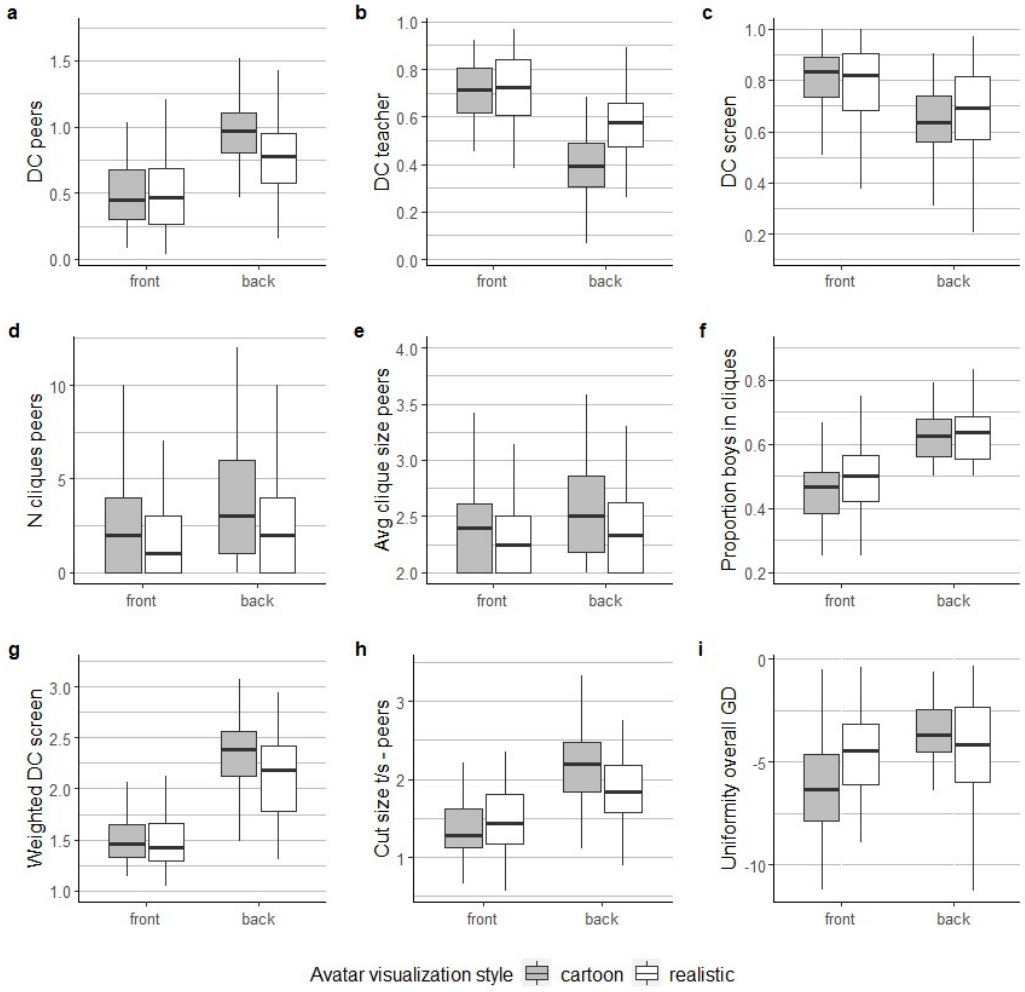
Interaction effects of social-related IVR classroom configurations on the structure of students' gaze-based attention networks

In addition to the main effects described above, we explored interaction effects of the different IVR configuration conditions on students' gaze-based attention networks. We expected that the effects of the avatars' visualization style and the variation in peers' performance-related behavior on students' gaze-based attention networks would be particularly pronounced when students were sitting in the back, where more peer learners were in the field of view (Hypothesis H1d). Whereas the effects of virtual peers' hand-raising behavior on students' gaze-based attention networks (see main effects in Section 4.1.1) were not affected by students' seating position, the results showed that the effects of the virtual avatar visualization style on students' gaze-based attention networks (see main effects in Section 4.1.2) were more pronounced when participants were sitting in the back of the IVR classroom. The results thus provided partial support for Hypothesis H1d.

As Figure C.13a shows, the focus of students' gaze transitions on virtual peer learners (i.e., degree centrality peers) visualized in a cartoon-style way was significantly greater in the back position; $F(1,258) = 11.87$, $p < .001$, $\eta_p^2 = 0.04$. Similarly, as can be seen in Figure C.13b, the degree centrality of the stylized teacher was significantly lower in the back; $F(1,258) = 19.77$, $p < .001$, $\eta_p^2 = 0.07$. There were no significant interaction effects of the IVR configuration conditions for connectedness of gazes (see Figure C.13d-e).

Moreover, as Figures C.13g-i show, the higher levels of the uniformity markers in the cartoon-style visualization of virtual avatars were particularly pronounced in the back position. The interaction effects were small compared to the main effects, but indicated statistically significant differences for the weighted degree centrality of the screen, $F(1,258) = 6.08$, $p = .014$, $\eta_p^2 = 0.02$, cut size, $F(1,258) = 12.50$, $p < .001$, $\eta_p^2 = 0.05$, and overall uniformity of gaze distribution, $F(1,258) = 21.44$, $p < .001$, $\eta_p^2 = 0.08$.

Figure C.13.: Boxplots of Structural Network Features by Seating Position and Avatar Visualization



Note. Values are averaged across hand-raising conditions. DC = Degree Centrality; N = Number; Avg = Average; t/s = teacher/screen; GD = Gaze Distribution.

How does the structure of students' gaze-based attention networks relate to their learning experiences in the IVR classroom?

The second research question asked how students' gaze-based attention networks in the IVR classroom relate to (a) their interest in the IVR lesson on computational thinking, (b) their evaluation of their own competence (i.e., situational self-concept) and (c) their performance on a test about the IVR lesson content afterwards. In this section, we report results of partial correlations between these outcomes and markers of students' gaze-based attention networks (i.e., structural variables describing gaze centrality, gaze connectedness among peers, and overall uniformity of gaze distribution), controlling for students' gender, general intelligence self-concept, and initial interest in the lesson topic as well as the IVR configuration conditions. We expected that more gaze centrality on peers and a higher connectedness of gazes among peers would be related to lower interest in the IVR lesson, lower situational self-concept, and lower test performance after the IVR lesson (Hypothesis H2a), and that more gaze centrality on the teacher and screen would be related to higher interest in the IVR lesson, higher situational self-concept, and better test performance after the IVR lesson (Hypothesis H2b). Table C.10 provides an overview of the results.

The results indicated that markers of students' gaze-based attention were most consistently related to their interest in the IVR lesson in the present study: Higher gaze centrality on the screen where the lesson content was presented was associated with higher interest in the IVR lesson topic ($\rho = 0.15$, $p = .011$, $d = 0.30$); conversely, the greater the gaze centrality on virtual peers, the lower the reported interest in the IVR lesson topic (Spearman's rho $\rho = -0.14$, $p = .021$, $d = 0.28$). Similarly, students' interest in the IVR lesson topic was negatively related to the number of observed cliques among peers ($\rho = -0.16$, $p = .009$, $d = 0.32$), the average clique size ($\rho = -0.17$, $p = .007$, $d = 0.35$) and the proportion of boys in the observed cliques ($\rho = -0.18$, $p = .003$, $d = 0.37$).

In addition, only one gaze-based feature each exhibited a small statistically significant relation with students' evaluation of their own competence in the IVR lesson (i.e., situational self-concept) and their performance on the posttest. The proportion of boys in the observed cliques was significantly related to students' situational self-concept: The more boys were in the observed cliques, the lower a student's situational self-concept ($\rho = -0.12$, $p = .043$, $d = 0.24$). In turn, degree centrality of the screen was positively related to students' performance on the posttest: The more students' gaze-based networks centered on the screen, the better their performance on the posttest ($\rho = 0.13$, $p = .028$, $d = 0.26$).

In sum, the results partly supported Hypotheses H2a and H2b.

Table C.9.: Descriptive Statistics for Structural Network Features in Different Hand-Raising Conditions

	20% (N = 72)		35% (N = 64)		65% (N = 60)		80% (N = 78)	
	M (SD)	*Mdn (MAD)*	*M (SD)*	*Mdn (MAD)*	*M (SD)*	*Mdn (MAD)*	*M (SD)*	*Mdn (MAD)*
DC peers	0.75 (0.35)	0.80 (0.35)	0.68 (0.32)	0.75 (0.35)	0.59 (0.32)	0.56 (0.33)	0.81 (0.36)	0.79 (0.39)
DC teacher	0.55 (0.19)	0.51 (0.20)	0.59 (0.20)	0.59 (0.22)	0.65 (0.18)	0.67 (0.20)	0.53 (0.20)	0.52 (0.21)
DC screen	0.70 (0.18)	0.72 (0.18)	0.73 (0.16)	0.73 (0.17)	0.76 (0.17)	0.82 (0.13)	0.66 (0.19)	0.68 (0.19)
N cliques peers	2.94 (3.06)	2.00 (2.97)	2.58 (2.72)	2.00 (2.97)	2.25 (2.45)	2.00 (2.97)	3.21 (2.66)	3.00 (2.97)
Avg clique size peers	2.31 (0.77)	2.32 (0.47)	2.33 (0.66)	2.38 (0.56)	2.21 (0.68)	2.33 (0.49)	2.37 (0.72)	2.44 (0.45)
Proportion boys in cliques	0.53 (0.20)	0.57 (0.11)	0.53 (0.17)	0.55 (0.11)	0.51 (0.20)	0.54 (0.18)	0.54 (0.19)	0.57 (0.14)
WDC screen	2.01 (0.54)	2.04 (0.69)	1.89 (0.57)	1.83 (0.67)	1.71 (0.45)	1.58 (0.40)	2.05 (0.52)	2.16 (0.59)
CS teacher/screen – peers	1.86 (0.57)	1.82 (0.64)	1.75 (0.63)	1.72 (0.59)	1.68 (0.55)	1.74 (0.57)	1.81 (0.52)	1.77 (0.65)
Uniformity overall GD	-4.28 (2.15)	-3.92 (2.00)	-4.48 (2.23)	-4.08 (2.04)	-4.84 (2.17)	-4.93 (2.22)	-4.58 (2.67)	-4.13 (2.81)

Note. Mean values and standard deviations (M and SD, respectively) as well as medians and median absolute deviations (Mdn and MAD, respectively) are reported. Values are averaged across the other configuration conditions. DC = Degree Centrality; N = Number; Avg = Average; WDC = Weighted Degree Centrality; CS = Cut Size; GD = Gaze Distribution

C.2.6. Discussion

The present study aimed to answer central questions about the configuration of IVR classrooms for educational research and practice and therefore examined how different socially relevant IVR classroom configuration features affect how students attend to different types of (social) information provided in the IVR classroom (RQ 1). We focused on three social-related IVR classroom configuration features that represent variations in spatial, visual, and behavioral proximity of the simulated virtual social counterparts, namely (a) students' positioning in the front vs. back of the IVR classroom, (b) the visualization style of virtual avatars as cartoonish vs. stylized, and (c) virtual peers' performance-related behavior in terms of different proportions of hand-raising students. Students' visual attention behavior was assessed via students' eye-tracking data, more specifically via features reflecting the structure of students' gaze-based attention networks (i.e., the gaze centrality on OOIs, connectedness of gazes among OOIs and uniformity of gaze distribution across OOIs). The results showed statistically significant differences between the social-related IVR classroom configuration conditions for all structural features of students' gaze-based attention networks in the classroom. To gain a more in-depth understanding of the structural features, in a second step, we examined how the structure of students' gaze-based attention networks relates to how students experienced the IVR classroom scenario (RQ 2). The results showed statistically significant relations to students' interest in the IVR lesson as well as their evaluation of their own competence (i.e., situational self-concept) and performance after the IVR lesson.

In the following sections, we discuss our findings in more detail. We focus on practical

conclusions concerning the configuration of IVR classrooms and the use of gaze-based attention networks to describe students' learning experiences in the IVR classroom in an objective and interpretable manner. However, we would like to highlight that these practical conclusions are naturally and closely intertwined with theoretical arguments. Theoretically speaking, our study showed that peer effects that have mostly been observed in "traditional" classrooms [383], [384] also play an important role in students' learning experiences in IVR classrooms. In this vein, our study extended established findings about peer effects in the classroom to the immersive virtual space and thereby provides the groundwork for many further studies that will examine how IVR classrooms can be used to advance educational research and practice.

Effects of social-related configurations on students' gaze-based attention networks in the IVR classroom

We examined students' gaze-based attention networks in an IVR classroom with different configuration conditions to obtain in-depth and objectively measurable insights into how different IVR classroom features affect how students attend to (social) information in the IVR classroom scenario. Taken together, our findings indicate that students' seating position as well as the visualization style and performance-related behavior of virtual avatars in an IVR classroom need to be carefully considered when using IVR for learning purposes or experimental classroom research. The present study's findings have important implications for educators and scholars aiming to select the best configuration to examine and utilize peer effects in an IVR classroom.

First, regarding *participants' position in the IVR classroom* (i.e., the spatial proximity of simulated social counterparts, such as the virtual teacher and virtual classmates), our findings indicated that positioning students in the front of an IVR classroom led to gaze-based attention networks that were more centered on the instructional content, whereas a position in the back was associated with a more comprehensive perception of all (social) information provided in the IVR classroom. Extending the findings of existing IVR studies suggesting better learning outcomes when sitting in the front of an IVR classroom [66], [91], our results provide evidence that, indeed, sitting in the front of an IVR classroom centers students' gaze more on the teacher and instructional content. Although this finding might be intuitive, considering that students sitting in the back have the whole class of peers in front of them, our results also indicated that students sitting in the back did not just focus more on peer learners, but distributed their attention more evenly across the classroom in general. More

Table C.10.: Partial Correlations of Gaze-Based Features with Interest in the Lesson, Situational Self-Concept, and Posttest Score

	Interest in IVR lesson	Situational self-concept	Posttest score
DC peers	-.14* [-0.26, -0.02]	-.10 [-0.21, 0.02]	-.12 [-0.24, 0.00]
DC teacher	.11 [-0.01, 0.23]	.07 [-0.05, 0.19]	.06 [-0.07, 0.17]
DC screen	.15* [0.04, 0.27]	.10 [-0.02, 0.21]	.13* [0.02, 0.25]
N cliques peers	-.16** [-0.27, -0.04]	-.06 [-0.18, 0.06]	-.04 [-0.16, 0.08]
Avg clique size peers	-.17** [-0.28, -0.05]	-.04 [-0.16, 0.08]	.00 [-0.12, 0.12]
Proportion boys in cliques	-.18** [-0.30, -0.06]	-.12* [-0.24, 0.00]	-.11 [-0.23, 0.01]
WDC screen	-.10 [-0.22, 0.02]	.00 [-0.12, 0.12]	-.03 [-0.15, 0.09]
CS teacher/screen – peers	.07 [-0.05, 0.19]	-.04 [-0.16, 0.08]	.01 [-0.11, 0.13]
Uniformity overall GD	.09 [-0.03, 0.21]	-.01 [-0.13, 0.11]	-.07 [-0.19, 0.05]

Note. Partial correlations controlling for gender, intelligence self-concept and initial interest in the lesson topic computational thinking as well as the IVR classroom configuration conditions. Variables are non-normally distributed; thus, Spearman's rho is reported. The Bonferroni correction was used to adjust for multiple significance tests. 95% confidence intervals are given in brackets. DC = Degree Centrality; N = Number; Avg = Average; WDC = Weighted Degree Centrality; CS = Cut Size; GD = Gaze Distribution. ** $p < .01$. * $p < .05$

balanced gaze transitions from the instructional content on the screen to different peers and more gaze transitions back and forth between the teacher or screen and peer learners rather than solely focusing on the instructional content might be an indication of more integrated visual attention toward (social) information in the classroom [106], [417]. Notably, in the present IVR lesson, the learning content was mainly provided by the virtual teacher and on the screen in the front of the classroom; therefore, students' attentional focus on the teacher and screen was desirable with regard to learning outcomes. However, given that peer learners can serve as an important source of information during instruction (e.g., [383]), a seating position in the back of the classroom might be more beneficial for learning in cases where virtual classmates are designed to be actively involved in the process of knowledge acquisition (e.g., as role models, to clarify misconceptions, etc.).

Second, regarding the *visualization style of the virtual avatars* (i.e., the visual proximity of simulated social counterparts, such as the virtual classmates), our findings indicate that for our sample of sixth graders, visualizing peer learners in a cartoon style was not just more cost- and time-efficient, but yielded no considerable disadvantages compared to a more realistic (i.e., stylized) visualization of peers: In fact, the students showed higher visual attention focus and gaze activity on cartoonish virtual peer learners, with particularly pronounced effects in the back seating position. Notably, alongside existing explanations for these findings (e.g., cartoonish peer learners are unusual and therefore attract more attention and cartoonish peers have larger head sizes which leads to increased visual attention; [3], [37]), the results of the present study point to an additional important aspect: When virtual avatars were visualized in cartoon style, we found (a) more equally distributed gazes between and screen and different peers and (b) more gaze transitions between instructional content (i.e., teacher and screen) and virtual peer learners, indicating that cartoon-style learners do not just attract attention to themselves, but are more engaging for students in an IVR classroom in general. This finding is not just important given that programming costs increase exponentially as virtual avatars become increasingly realistic, it also points to potential affordances of cartoonish characters when aiming to design IVR classroom environments that invite high engagement with virtual avatars (e.g., in collaborative learning scenarios or with virtual classmates as emotional support).

Third, with regard to *peer learners' performance-related behavior* (i.e., the behavioral proximity of simulated virtual classmates), our findings indicated that virtual peer learners' hand-raising had the greatest effect on students' visual attention distribution in the IVR classroom when it was most salient and unambiguous (i.e., a clear minority or majority of

peers raising their hands). Against our expectation that more hand-raising would lead to more gaze centrality on peers (and respectively less on the teacher and screen), we found the highest gaze centrality on peers in the 'extreme' conditions of 20% and 80% hand-raising (and respectively the highest gaze centrality on the teacher and screen in the more moderate conditions of 35% and 65% hand-raising). Importantly, these effects were not affected by seating position, suggesting that the most salient hand-raising conditions of 20% and 80% were recognized most by students regardless of whether they were positioned in the front or the back of the IVR classroom. Indicating that 'social manipulations' in an IVR classroom are particularly effective when they are very clearly interpretable (i.e., almost none or pretty much all peers are raising their hands), this finding has important implications for the design of peer learners' behavior in IVR classrooms in both educational research and practice. More specifically, this finding suggests that peers' (performance-related) behavior needs to be configured to be as unambiguous as possible (a) to investigate respective effects of peer behavior on student learning, and (b) to strategically deploy respective behaviors in the design of simulated virtual peer learners as pedagogical agents in IVR classroom-based learning applications (see, e.g., [91], [386], [406]).

Using gaze-based attention networks to gain insights into students' learning experiences in the IVR classroom

Given that the use of graph-based analysis is a relatively new approach for analyzing gaze data and visual attention, especially in an IVR classroom setting and in relation to students' learning experiences, we were interested in how the structure of students' gaze-based attention networks relates to central outcome variables in the context of classroom learning (i.e., students' interest, situational self-concept and performance).

In line with our expectations, we found significant relations between students' learning experiences in the IVR classroom and their gaze centrality on peers, the teacher and the screen as well as with the connectedness of gazes among peers. Notably, the examined structural features of students' gaze-based attention networks allowed us to capture specific aspects of students' visual attention distribution, such as gaze centrality on certain objects of interest or visual attention focus on different subgroups (e.g., the proportion of boys in the observed cliques). In the end, we found relations between educational outcomes and the structure of students' gaze-based attention networks exclusively for features describing visual attention tied to different objects of interest (e.g., degree centrality on the screen, proportion of boys in observed cliques), whereas more general descriptions of students' gaze behavior

(i.e., markers of uniformity) were not related to any of the examined educational outcomes.

As expected, the more interested students were in the IVR lesson content, the more they focused on the instructional content and the less they attended to social information provided by their peers. Accordingly, students' test performance after the IVR lesson was positively related to their visual attention focus on the screen. Given that everything necessary to obtain a good test score was presented on the screen, this finding is in line with our expectations. At the same time, considering that the most important content of the IVR lesson was also presented orally by the teacher and the audio was the same in all IVR configurations, it is not surprising that the effect of visual attention on the lesson content was comparably small. Notably, students' performance after the IVR lesson was not related to their visual attention on peers. On the one hand, this finding might be considered reassuring given the potential detrimental effect of peers as distractions from the instructional content; on the other hand, considering the effect of only the 'extreme' hand-raising behaviors on students' gaze-based attention networks, this finding additionally highlights that the manipulation of hand-raising behavior in the present study was not 'powerful' enough to make use of potential beneficial effects of peers—for instance, as pedagogical agents. Hence, it might be worthwhile to implement fewer but very salient peer avatars (in line with suggestions by [301]).

Regarding students' situational self-concept, only the proportion of boys in the observed cliques exhibited a negative relation to how students evaluated their own competence during the IVR lesson. In line with common assumptions in research on reference group effects, we expected similar findings for the degree centrality of peer learners as well as for the number and average size of cliques among virtual peers. Hence, the present study's finding highlights the role of very specific social information for students' self-evaluations. Although the observed effect is small and needs further investigation in future studies, we argue that this result is particularly interesting given that our IVR lesson concerned the topic of computational thinking, which might be associated with gender stereotypes that affect who students compare themselves to and how they consequently evaluate themselves [433]–[436].

Speaking to methodological contributions regarding the use of gaze-based attention networks, we would like to highlight that the successful application of respective algorithms is highly context- and task-specific. The basis for our gaze-based attention networks are adjacency matrices, which other studies typically use in scanpath analyses or as input for Support Vector Classification or other machine learning algorithms [114]. The purpose of such methods is to incorporate the spatiotemporal, sequential nature of eye movements

into the analyses [437]. In our study, we created networks from gaze transitions in an IVR classroom to extract features that describe students' visual attention. Attesting to the high context-specificity of the algorithms used, we argue that the possibilities and limitations of this approach need to be carefully tested and adapted when applied in other scenarios. However, we see two main advantages of using and further developing this methodology. First, most of the current scanpath algorithms make it difficult to explain results because their features are highly abstract. Using graph or network theory instead makes eye movement patterns visible and extracted features comprehensible. Second, combining methodological approaches that so far have only been sparsely connected (i.e., analysis of eye-tracking data from IVR environments and graph/network theory) offers the possibility of developing a multidisciplinary perspective on the nature of human visual attention.

Limitations and future directions

In the present study, we applied graph theory and network analysis to students' eye-tracking data from an IVR lesson to examine differences in students' gaze-based attention networks in an IVR classroom with different configurations. We manipulated three central configuration features of the IVR classroom (i.e., students' position, the visualization style of virtual avatars, and performance-related behavior of virtual peer learners) and examined central structural variables describing students' gaze-based attention networks in three categories (i.e., gaze centrality, connectedness of gazes among peer learners, and uniformity of gaze distribution across OOIs). Notably, although our approach yielded a number of important findings, we would also like to point out some limitations that provide great potential for future research regarding the configurations of IVR classroom environments and students' individual responses to them.

In terms of IVR classroom design, we would like to highlight four critical aspects. First, we applied a neutral design of the classroom environment and focused on specificities of virtual avatar design as a particularly socially relevant feature. However, it should be noted that the overall IVR classroom design (e.g., wall color, posters, lighting, etc.) provides many additional opportunities to further guide students' attentional focus and affect their perception of the IVR classroom scenario (see, e.g., [438]). Second, whereas we found significant positive effects of the cartoonish avatar visualization on students' overall engagement with virtual peer learners, it needs to be considered that our sample consisted of sixth graders who might have been more engaged with cartoon learners in comparison to adults or older students. Future research should extend these findings and examine the effects in different

age groups. Third, in order to render 24 virtual peer agents, we needed to keep the visual realism under a certain level; hence, even our "more realistic" (i.e., stylized) avatars do not represent the highest degree of realism that is currently possible. In addition, we only varied the visualization style of the virtual avatars; both visualization styles were based on the same motion captures and therefore, the avatars' movements and gestures were identical across the different visualizations. In light of previous work demonstrating the importance of a good match between behavioral and photographic realism [439], [440], it seems worthwhile for future research to further explore different visualization styles of peer learners in combination with different simulations of performance-related behaviors. Fourth, we only varied the appearance of the virtual avatars of the peer learners and the teacher, whereas the participating students were not represented by an avatar in our IVR classroom. In light of the substantive body of research examining the effects of self-representation via avatars on users' behavior and experience in IVR (see meta-analysis by [441]), implementing representations of participating students in the IVR classroom seems worth investigating further.

With regard to the virtual peer learners' behavior, we manipulated their performance-related behavior via hand-raising as an indicator of their performance and overall behavioral engagement. Based on the present study's findings suggesting (a) an effect of peers' hand-raising on students' gaze-based attention networks in the IVR classroom and (b) a relation between gaze-based attentional focus on peers and central learning outcomes, we argue that future research should extend this line of research and consider additional variations of peer behavior and classroom composition. In addition, given that the proportion of observed boys also had an effect on students' self-evaluations, future research should make use of the affordances of IVR to examine gender differences with regard to peer effects (see, e.g., Chang et al., 2019; Lee et al., 2014).

Moreover, our IVR was fully preprogrammed, which allowed for maximum standardization and therefore systematic insights into students' IVR experience. However, we believe that the implementation of some interaction options for participating students might provide additional valuable insights into reference group effects in an interactive yet standardized setting. For instance, Liao et al. [301] demonstrated the impact of virtual classmates on students' learning by implementing interactive virtual classmates with time-anchored comments and behaviors based on content and valence analyses of participants' prior comments during instruction. Combining such approaches with analyses of students' actual gaze-based attentional networks in the classroom seems like a promising avenue to gain insights into (a)

how students make use of (social) information provided in the IVR classroom and (b) how the ideal IVR classroom for student learning should consequently be configured.

Notably, when students learn with new technologies, such as IVR, the pure novelty and unfamiliarity of the technological learning environment can affect their learning experiences and respective outcomes [442], [443]. In the present study, about half of the participants used IVR technology for the first time; importantly, participants perceived the IVR classroom scenario as very similar to a real-world classroom that they are familiar with. Nevertheless, we would like to highlight that the effect of novelty and unfamiliarity is important to consider for future studies when examining technology-enhanced learning environments, such as IVR classrooms, and students' learning experiences in them.

Lastly, our IVR lesson lasted only 15 minutes, and we aggregated students' gaze-based attention networks over the entire lesson period in order to gain insights into their visual attention toward (social) information provided in the IVR classroom. Whereas our approach yielded important insights into the effects of different IVR configurations (see Section 5.1) and the use of gaze-based attention networks to gain insights into students' visual attention and learning experiences in the IVR classroom (see Section 5.2), we argue that future research should see whether our findings replicate in other and longer IVR classroom scenarios. Moreover, the graph-based analysis of gaze data has great potential for additionally examining dynamics and changes in students' gaze-based attention networks (see, e.g., [367]), such as whether attentional focus on the teacher decreases over time or whether students focus their gaze on certain students at important conversational points. In addition, we encourage future research to explore integrating behavioral information from hand or head movements into corresponding analyses (see for the potential of head movements, e.g., [444], [445]). Such studies might yield additional valuable insights into students' visual attention toward (social) information in the IVR classroom during different phases of instruction.

C.2.7. Conclusion

Aiming to examine and utilize peer effects in an IVR classroom, the present study answers central questions about the effects of social-related IVR classroom configurations on students' visual attention and learning experiences with a full class of simulated virtual peer learners. Overall, our results underline the potential of transforming "traditional" classrooms into immersive virtual reality scenarios for research purposes and effective learning scenarios. With regard to social-related IVR classroom configuration, the present study's findings indicate that the positioning of students in the IVR classroom, the visualization style of virtual avatars,

as well as the performance-related behavior of virtual peer learners are decisive features to consider when configuring an IVR classroom. By examining students' gaze-based attention networks during instruction in an IVR classroom, we were able to gain valuable insights into the effects of different socially relevant IVR classroom configurations on students' perception of the IVR classroom environment and visual attention toward respective (social) information. Both educational researchers and practitioners are encouraged to carefully consider potential (side) effects of different social-related IVR classroom configurations in light of their individual intentions and (research or learning) goals for using an IVR classroom.

Author note We posted all data and data analysis scripts on the Open Science Framework (OSF) under the following link: `https://osf.io/pek4q/?view_only=ef151fd06ac8413a827020d4264b3c8d`.
Declaration of competing interest We have no conflicts of interest to disclose.

C.2.8. Acknowledgements

This research was supported by a grant to Richard G¨ollner funded by the Ministry of Science, Research and the Arts of the state of Baden-Württemberg and the University of Tübingen as part of the Promotion Program for Junior Researchers. Lisa Hasenbein and Philipp Stark are doctoral candidates at the LEAD Graduate School & Research Network, which is funded by the Ministry of Science, Research and the Arts of the state of Baden-Württemberg within the sustainability funding framework for projects within the Excellence Initiative II. We would like to thank Stephan Soller, Sandra Hahn and Sophie Fink from the Institute for Games, Department of Computer Science and Media, at the Hochschule der Medien Stuttgart for their extensive work preparing the immersive virtual reality classroom used in this study. Our thanks also go to Luzia Leifheit for providing the learning materials for the Computational Thinking course used in the present study.

Bibliography

[1] P. Stark, E. Bozkir, W. Sójka, M. Huff, E. Kasneci, and R. Göllner, "The impact of presentation modes on mental rotation processing: A comparative analysis of eye movements and performance", *Scientific Reports*, 2024. DOI: 10.1038/s41598-024-60370-6.

[2] P. Stark, A. Tobias, O. Milo, and K. Enkelejda, "Pupil diameter during counting tasks as potential baseline for virtual reality experiments", in *2023 Symposium on Eye Tracking Research and Applications (ETRA '23)*, Germany: ACM, Jun. 30, 2023, p. 7. DOI: 10.1145/3588015.3588414.

[3] E. Bozkir, P. Stark, H. Gao, L. Hasenbein, J.-U. Hahn, E. Kasneci, and R. Göllner, "Exploiting object-of-interest information to understand attention in VR classrooms", in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, Mar. 2021, pp. 597–605. DOI: 10.1109/VR50410.2021.00085.

[4] P. Stark, A. Jung, J.-U. Hahn, E. Kasneci, and R. Göllner, "Using gaze transition entropy to detect classroom discourse in a virtual reality classroom", in *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications (ETRA '24)*, Glasgow, UK: ACM, 2024. DOI: 10.1145/3649902.3653335.

[5] P. Stark, L. Hasenbein, E. Kasneci, and R. Göllner, "Gaze-based attention network analysis in a virtual reality classroom", *MethodsX*, vol. 12, p. 102 662, Jun. 1, 2024. DOI: 10.1016/j.mex.2024.102662.

[6] L. Hasenbein, P. Stark, U. Trautwein, A. C. M. Queiroz, J. Bailenson, J.-U. Hahn, and R. Göllner, "Learning with simulated virtual classmates: Effects of social-related configurations on students' visual attention and learning experiences in an immersive virtual reality classroom", *Computers in Human Behavior*, vol. 133, p. 107 282, Aug. 1, 2022. DOI: 10.1016/j.chb.2022.107282.

[7] J. Ferdinand, H. Gao, P. Stark, E. Bozkir, J.-U. Hahn, E. Kasneci, and R. Göllner, "The impact of a usefulness intervention on students' learning achievement in a virtual biology lesson: An eye-tracking-based approach", *Learning and Instruction*, vol. 90, p. 101 867, Apr. 1, 2024. DOI: 10.1016/j.learninstruc.2023.101867.

[8] H. Gao, E. Bozkir, P. Stark, P. Goldberg, G. Meixner, E. Kasneci, and R. Göllner, "Detecting teacher expertise in an immersive VR classroom: Leveraging fused sensor data with explainable machine learning models", in *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Oct. 2023, pp. 683–692. DOI: 10.1109/ISMAR59233.2023.00083.

[9] L. Hasenbein, P. Stark, T. Ulrich, H. Gao, E. Kasneci, and R. Göllner, "Investigating social comparison behaviour in an immersive virtual reality classroom based on eye-movement data", *Scientific Reports*, vol. 13, no. 1, p. 14 672, Sep. 6, 2023. DOI: 10.1038/s41598-023-41704-2.

[10] M. Slater and M. V. Sanchez-Vives, "Enhancing our lives with immersive virtual reality", *Frontiers in Robotics and AI*, vol. 3, p. 74, 2016. DOI: 10.3389/frobt.2016.00074.

[11] Y. Huang, E. Richter, T. Kleickmann, and D. Richter, *Virtual reality in teacher education from 2010 to 2020: A review of program implementation, intended outcomes, and effectiveness measures*, Dec. 1, 2021. DOI: 10.35542/osf.io/ye6uw.

[12] G. Riva, "Virtual reality in clinical psychology", *Reference Module in Neuroscience and Biobehavioral Psychology*, pp. 91–105, 2022. DOI: 10.1016/B978-0-12-818697-8.00006-6.

[13] S. de la Rosa and M. Breidt, "Virtual reality: A new track in psychological research", *British Journal of Psychology*, vol. 109, no. 3, pp. 427–430, 2018. DOI: 10.1111/bjop.12302.

[14] M. Vasser and J. Aru, "Guidelines for immersive virtual reality in psychological research", *Current Opinion in Psychology*, vol. 36, pp. 71–76, Dec. 1, 2020. DOI: 10.1016/j.copsyc.2020.04.010.

[15] J. O. Bailey and J. N. Bailenson, "Considering virtual reality in children's lives", *Journal of Children and Media*, vol. 11, no. 1, pp. 107–113, 2017. DOI: 10.1080/17482798.2016.1268779.

[16] V. Clay, P. König, and S. König, "Eye tracking in virtual reality", *Journal of Eye Movement Research*, vol. 12, no. 1, 2019. DOI: 10.16910/jemr.12.1.3.

[17] T. D. Parsons, "Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences", *Frontiers in Human Neuroscience*, vol. 9, Dec. 11, 2015. DOI: 10.3389/fnhum.2015.00660.

[18] O. D. Kothgassner and A. Felnhofer, "Does virtual reality help to cut the gordian knot between ecological validity and experimental control?", *Annals of the International Communication Association*, vol. 44, no. 3, pp. 210–218, Jul. 2, 2020. DOI: 10.1080/23808985.2020.1792790.

[19] I. Wohlgenannt, A. Simons, and S. Stieglitz, "Virtual reality", *Business & Information Systems Engineering*, vol. 62, no. 5, pp. 455–461, Oct. 1, 2020. DOI: 10.1007/s12599-020-00658-9.

[20] D. Bores-García, R. Cano-de-la-Cuerda, M. Espada, N. Romero-Parra, D. Fernández-Vázquez, J. M. Delfa-De-La-Morena, V. Navarro-López, and D. Palacios-Ceña, "Educational research on the use of virtual reality combined with a practice teaching style in physical education: A qualitative study from the perspective of researchers", *Education Sciences*, vol. 14, no. 3, p. 291, Mar. 2024. DOI: 10.3390/educsci14030291.

[21] G. Billingsley, S. Smith, S. Smith, and J. Meritt, "A systematic literature review of using immersive virtual reality technology in teacher education", *Journal of Interactive Learning Research*, vol. 30, no. 1, pp. 65–90, 2019.

[22] L. Jensen and F. Konradsen, "A review of the use of virtual reality head-mounted displays in education and training", *Education and Information Technologies*, vol. 23, no. 4, pp. 1515–1529, Jul. 1, 2018. DOI: 10.1007/s10639-017-9676-0.

[23] G. Fauville, A. Voşki, M. Mado, J. N. Bailenson, and A. Lantz-Andersson, "Underwater virtual reality for marine education and ocean literacy: Technological and psychological potentials", *Environmental Education Research*, pp. 1–25, Mar. 14, 2024. DOI: 10.1080/13504622.2024.2326446.

[24] J. Ferdinand, S. Soller, J.-U. Hahn, J. Parong, and R. Göllner, "Enhancing the effectiveness of virtual reality in science education through an experimental intervention involving students' perceived usefulness of virtual reality", *Technology, Mind, and Behavior*, vol. 4, no. 1, Feb. 13, 2023. DOI: 10.1037/tmb0000084.

[25] D. Allison and L. F. Hodges, "Virtual reality for education?", in *Proceedings of the ACM symposium on Virtual reality software and technology*, ser. VRST '00, New York, NY, USA: ACM, 2000, pp. 160–165. DOI: 10.1145/502390.502420.

[26] D. Kamińska, T. Sapiński, S. Wiak, T. Tikk, R. Haamer, E. Avots, A. Helmi, C. Ozcinar, and G. Anbarjafari, "Virtual reality and its applications in education: Survey", *Information (Switzerland)*, vol. 10, p. 318, 2019. DOI: 10.3390/info10100318.

[27] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda", *Computers & Education*, vol. 147, p. 103 778, Apr. 1, 2020. DOI: 10.1016/j.compedu.2019.103778.

[28] E. Johnston, G. Olivas, P. Steele, C. Smith, and L. Bailey, "Exploring pedagogical foundations of existing virtual reality educational applications: A content analysis study", *Journal of Educational Technology Systems*, vol. 46, no. 4, pp. 414–439, 2017. DOI: 10.1177/0047239517745560.

[29] R. Ristor, S. Morélot, A. Garrigou, and B. N' Kaoua, "Virtual reality for fire safety training: Study of factors involved in immersive learning", *Virtual Reality*, vol. 27, pp. 2237–2254, May 8, 2023. DOI: 10.1007/s10055-022-00743-2.

[30] N. D. Vega, R. Rahayu, and N. Basri, "The role of virtual reality in enhancing the effectiveness of teaching english for specific purposes", *Research and Innovation in Applied Linguistics [RIAL]*, vol. 2, no. 1, pp. 1–13, Feb. 29, 2024. DOI: 10.31963/rial.v2i1.4386.

[31] D. Allcoat and A. v. Mühlenen, "Learning in virtual reality: Effects on performance, emotion and engagement", *Research in Learning Technology*, vol. 26, Nov. 27, 2018. DOI: 10.25304/rlt.v26.2140.

[32] J. Parong and R. E. Mayer, "Learning science in immersive virtual reality", *Journal of Educational Psychology*, vol. 110, no. 6, pp. 785–797, 2018. DOI: 10.1037/edu0000241.

[33] J. Zhao, T. Sensibaugh, B. Bodenheimer, T. McNamara, A. Nazareth, N. Newcombe, M. Minear, and A. Klippel, "Desktop versus immersive virtual environments: Effects on spatial learning", *Spatial Cognition & Computation*, vol. 3, pp. 1–36, Sep. 2020. DOI: 10.1080/13875868.2020.1817925.

[34] A. P. Lawson and R. E. Mayer, "Effect of pre-training and role of working memory characteristics in learning with immersive virtual reality", *International Journal of Human–Computer Interaction*, pp. 1–18, 2024. DOI: 10.1080/10447318.2024.2325176.

[35] H. Choi, J. Kwon, and S. Nam, "Research on the application of gaze visualization interface on virtual reality training systems", *Journal on Multimodal User Interfaces*, vol. 17, no. 3, pp. 203–211, Sep. 1, 2023. DOI: 10.1007/s12193-023-00409-6.

[36] M. A. Rau and T. Herder, "Under which conditions are physical versus virtual representations effective? contrasting conceptual and embodied mechanisms of learning", *Journal of Educational Psychology*, vol. 113, no. 8, pp. 1565–1586, 2021. DOI: 10.1037/edu0000689.

[37] H. Gao, E. Bozkir, L. Hasenbein, J.-U. Hahn, R. Göllner, and E. Kasneci, "Digital transformations of classrooms in virtual reality", in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 483, New York, NY, USA: ACM, May 6, 2021, pp. 1–10.

[38] Y. Han, Y. Miao, J. Lu, M. Guo, and Y. Xiao, "Exploring intervention strategies for distracted students in VR classrooms", in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, ser. CHI EA '22, New York, NY, USA: ACM, Apr. 27, 2022, pp. 1–7. DOI: 10.1145/3491101.3519627.

[39] L. Hasenbein, U. Trautwein, J.-U. Hahn, S. Soller, and R. Göllner, "An experimental test of the big-fish-little-pond effect using an immersive virtual reality classroom", *Instructional Science*, 2023. DOI: 10.1007/s11251-023-09646-4.

[40] B. Coleman, S. Marion, A. Rizzo, J. Turnbull, and A. Nolty, "Virtual reality assessment of classroom – related attention: An ecologically relevant approach to evaluating the effectiveness of working memory training", *Frontiers in Psychology*, vol. 10, p. 1851, 2019. DOI: https://doi.org/10.3389/fpsyg.2019.01851.

[41] A. C. Roberts, Y. W. Yeap, H. S. Seah, E. Chan, C.-K. Soh, and G. I. Christopoulos, "Assessing the suitability of virtual reality for psychological testing", *Psychological Assessment*, vol. 31, no. 3, pp. 318–328, 2019. DOI: 10.1037/pas0000663.

[42] T. D. Parsons and A. S. Phillips, "Virtual reality for psychological assessment in clinical practice", *Practice Innovations*, vol. 1, pp. 197–217, 2016. DOI: 10.1037/pri0000028.

[43] Y. Huang, E. Richter, T. Kleickmann, and D. Richter, "Class size affects preservice teachers' physiological and psychological stress reactions: An experiment in a virtual reality classroom", *Computers & Education*, vol. 184, p. 104 503, Jul. 1, 2022. DOI: 10.1016/j.compedu.2022.104503.

[44] J. Brookes, M. Warburton, M. Alghadier, M. Mon-Williams, and F. Mushtaq, "Studying human behavior with virtual reality: The unity experiment framework", *Behavior Research Methods*, vol. 52, no. 2, pp. 455–463, Apr. 1, 2020. DOI: 10.3758/s13428-019-01242-0.

[45] H. Gao, L. Hasenbein, E. Bozkir, R. Göllner, and E. Kasneci, "Exploring gender differences in computational thinking learning in a VR classroom: Developing machine learning models using eye-tracking data and explaining the models", *International Journal of Artificial Intelligence in Education*, vol. 33, no. 4, pp. 929–954, Dec. 1, 2023. DOI: 10.1007/s40593-022-00316-z.

[46] S. Eftekharifar, A. Thaler, A. O. Bebko, and N. F. Troje, "The role of binocular disparity and active motion parallax in cybersickness", *Experimental Brain Research*, vol. 239, no. 8, pp. 2649–2660, Aug. 2021. DOI: 10.1007/s00221-021-06124-6.

[47] X. Wang and N. Troje, "Relating visual and pictorial space: Binocular disparity for distance, motion parallax for direction", *Visual Cognition*, vol. 31, pp. 1–19, Apr. 24, 2023. DOI: 10.1080/13506285.2023.2203528.

[48] M. I. Berkman and E. Akan, "Presence and immersion in virtual reality", in *Encyclopedia of Computer Graphics and Games*, N. Lee, Ed., Cham: Springer International Publishing, 2019, pp. 1–10.

[49] F. Biocca, C. Harms, and J. K. Burgoon, "Toward a more robust theory and measure of social presence: Review and suggested criteria", *Presence: Teleoperators and Virtual Environments*, vol. 12, no. 5, pp. 456–480, Oct. 1, 2003. DOI: 10.1162/105474603322761270.

[50] H. Gao, L. Frommelt, and E. Kasneci, "The evaluation of gait-free locomotion methods with eye movement in virtual reality", in *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, Oct. 2022, pp. 530–535. DOI: 10.1109/ISMAR-Adjunct57072.2022.00112.

[51] H. Gao and E. Kasneci, "Eye-tracking-based prediction of user experience in VR locomotion using machine learning", *Computer Graphics Forum*, vol. 41, no. 7, pp. 589–599, 2022. DOI: 10.1111/cgf.14703.

[52] X. Pan and A. F. d. C. Hamilton, "Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape", *British Journal of Psychology*, vol. 109, no. 3, pp. 395–417, 2018. DOI: 10.1111/bjop.12290.

[53] H. Gao, L. Hasenbein, E. Bozkir, R. Göllner, and E. Kasneci, "Evaluating the effects of virtual human animation on students in an immersive VR classroom using eye movements", in *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '22, New York, NY, USA: ACM, Nov. 29, 2022, pp. 1–11. DOI: 10.1145/3562939.3565623.

[54] J. Blascovich, A. Beall, K. Swinth, C. Hoyt, and J. Bailenson, "Immersive virtual environment technology as a methodological tool for social psychology", *Psychol. Inq*, vol. 13, no. 2, pp. 103–124, 2002. DOI: 10.1207/S15327965PLI1302_01.

[55] H. E. Yaremych and S. Persky, "Tracing physical behavior in virtual reality: A narrative review of applications to social psychology", *Journal of Experimental Social Psychology*, vol. 85, p. 103 845, Nov. 1, 2019. DOI: 10.1016/j.jesp.2019.103845.

[56] K. Holmqvist, M. Nyström, and F. Mulvey, "Eye tracker data quality: What it is and how to measure it", in *Proceedings of the Symposium on Eye Tracking Research and Applications*, ser. ETRA '12, New York, NY, USA: ACM, 2012, pp. 45–52. DOI: 10.1145/2168556.2168563.

[57] M. K. Eckstein, B. Guerra-Carrillo, A. T. Miller Singley, and S. A. Bunge, "Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development?", *Developmental Cognitive Neuroscience*, vol. 25, pp. 69–91, 2017. DOI: 10.1016/j.dcn.2016.11.001.

[58] I. B. Adhanom, P. MacNeilage, and E. Folmer, "Eye tracking in virtual reality: A broad review of applications and challenges", *Virtual Reality*, vol. 27, no. 2, pp. 1481–1505, Jan. 18, 2023. DOI: 10.1007/s10055-022-00738-z.

[59] E. Bozkir, S. Özdel, M. Wang, B. David-John, H. Gao, K. Butler, E. Jain, and E. Kasneci, *Eye-tracked virtual reality: A comprehensive survey on methods and privacy challenges*, May 23, 2023. DOI: 10.48550/arXiv.2305.14080.

[60] E. Kasneci, H. Gao, S. Ozdel, V. Maquiling, E. Thaqi, C. Lau, Y. Rong, G. Kasneci, and E. Bozkir, *Introduction to eye tracking: A hands-on tutorial for students and practitioners*, Apr. 1, 2024. DOI: 10.48550/arXiv.2404.15435.

[61] J. K. Kaakinen, "What can eye movements tell us about visual perception processes in classroom contexts? commentary on a special issue", *Educational Psychology Review*, vol. 33, no. 1, pp. 169–179, Mar. 1, 2021. DOI: 10.1007/s10648-020-09573-7.

[62] M.-L. Lai, M.-J. Tsai, F.-Y. Yang, C.-Y. Hsu, T.-C. Liu, S. W.-Y. Lee, M.-H. Lee, G.-L. Chiou, J.-C. Liang, and C.-C. Tsai, "A review of using eye-tracking technology in exploring learning from 2000 to 2012", *Educational Research Review*, vol. 10, pp. 90–115, Dec. 1, 2013. DOI: 10.1016/j.edurev.2013.10.001.

[63] S. Rapti, T. Sapounidis, and S. Tselegkaridis, "Investigating educators' and students' perspectives on virtual reality enhanced teaching in preschool", *Early Childhood Education Journal*, Apr. 5, 2024. DOI: 10.1007/s10643-024-01659-z.

[64] T. Shebane and J. Dehinbo, "Developing and exploring the use of virtual reality learning system to teach mathematics toward minimizing failure rate", *Journal of Computers in Mathematics and Science Teaching*, vol. 39, no. 2, pp. 125–147, Apr. 2020.

[65] E. Han and J. N. Bailenson, "Lessons for/in virtual classrooms: Designing a model for classrooms inside virtual reality", *Communication Education*, vol. 73, no. 2, pp. 234–243, Apr. 2, 2024. DOI: 10.1080/03634523.2024.2312879.

[66] F. Blume, R. Göllner, K. Moeller, T. Dresler, A.-C. Ehlis, and C. Gawrilow, "Do students learn better when seated close to the teacher? a virtual classroom study considering individual levels of inattention and hyperactivity-impulsivity", *Learning and Instruction*, vol. 61, pp. 138–147, 2019. DOI: 10.1016/j.learninstruc.2018.10.004.

[67] S.-H. Seo, E. Kim, P. Mundy, J. Heo, and K. K. Kim, "Joint attention virtual classroom: A preliminary study", *Psychiatry Investigation*, vol. 16, no. 4, pp. 292–299, 2019. DOI: 10.30773/pi.2019.02.08.

[68] F. L. Kooi and A. Toet, "Visual comfort of binocular and 3d displays", *Displays*, vol. 25, no. 2, pp. 99–108, Aug. 1, 2004. DOI: 10.1016/j.displa.2004.07.004.

[69] B. J. Concannon, S. Esmail, and M. Roduta Roberts, "Head-mounted display virtual reality in post-secondary education and skill training", *Frontiers in Education*, vol. 4, Aug. 14, 2019. DOI: 10.3389/feduc.2019.00080.

[70] J. C. Almenara and J. B. Osuna, "The educational possibilities of augmented reality", *Journal of New Approaches in Educational Research*, vol. 5, no. 1, pp. 44–50, Jan. 15, 2016. DOI: 10.7821/naer.2016.1.140.

[71] S. Bjork and J. Holopainen, *Patterns in game design* (Charles river media game development series). Charles River Media, 2005.

[72] U. Korisky and L. Mudrik, "Dimensions of perception: 3d real-life objects are more readily detected than their 2d images", *Psychological Science*, vol. 32, no. 10, pp. 1636–1648, Oct. 1, 2021. DOI: 10.1177/09567976211010718.

[73] M. Lamb, M. Brundin, E. Perez Luque, and E. Billing, "Eye-tracking beyond peripersonal space in virtual reality: Validation and best practices", *Frontiers in Virtual Reality*, vol. 3, 2022. DOI: https://doi.org/10.3389/frvir.2022.864653.

[74] C. Snelson and Y.-C. Hsu, "Educational 360-degree videos in virtual reality: A scoping review of the emerging research", *TechTrends*, vol. 64, no. 3, pp. 404–412, 2020. DOI: 10.1007/s11528-019-00474-3.

[75] P. Ugwitz, O. Kvarda, Z. Juříková, Č. Šašinka, and S. Tamm, "Eye-tracking in interactive virtual environments: Implementation and evaluation", *Applied Sciences*, vol. 12, no. 3, p. 1027, Jan. 2022. DOI: 10.3390/app12031027.

[76] A. A. Rizzo, J. G. Buckwalter, T. Bowerly, C. Van Der Zaag, L. Humphrey, U. Neumann, C. Chua, C. Kyriakakis, A. Van Rooyen, and D. Sisemore, "The virtual classroom: A virtual reality environment for the assessment and rehabilitation of attention deficits", *CyberPsychology & Behavior*, vol. 3, no. 3, pp. 483–499, 2000. DOI: 10.1089/10949310050078940.

[77] Z. Tang, X. Liu, H. Huo, M. Tang, X. Qiao, D. Chen, Y. Dong, L. Fan, J. Wang, X. Du, J. Guo, S. Tian, and Y. Fan, "Eye movement characteristics in a mental rotation task presented in virtual reality", *Frontiers in Neuroscience*, vol. 17, 2023. DOI: 10.3389/fnins.2023.1143006.

[78] M. Slater and S. Wilbur, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments", *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 6, pp. 603–616, Dec. 1, 1997. DOI: 10.1162/pres.1997.6.6.603.

[79] T. Schubert, F. Friedmann, and H. Regenbrecht, "The experience of presence: Factor analytic insights", *Presence*, vol. 10, no. 3, pp. 266–281, 2001. DOI: 10.1162/105474601300343603.

[80] G. Makransky and G. B. Petersen, "The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality", *Educational Psychology Review*, vol. 33, no. 3, pp. 937–958, 2021. DOI: 10.1007/s10648-020-09586-2.

[81] E.-L. Sallnäs, K. Rassmus-Gröhn, and C. Sjöström, "Supporting presence in collaborative environments by haptic force feedback", *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 4, pp. 461–476, 2000. DOI: 10.1145/365058.365086.

[82] C. S. Oh, J. N. Bailenson, and G. F. Welch, "A systematic review of social presence: Definition, antecedents, and implications", *Frontiers in Robotics and AI*, vol. 5, Oct. 15, 2018. DOI: 10.3389/frobt.2018.00114.

[83] C. Schwartz, G. Bente, A. Gawronski, L. Schilbach, and K. Vogeley, "Responses to nonverbal behaviour of dynamic virtual characters in high-functioning autism", *Journal of Autism and Developmental Disorders*, vol. 40, no. 1, pp. 100–111, Jan. 1, 2010. DOI: 10.1007/s10803-009-0843-z.

[84] M. Tomasello, M. Carpenter, J. Call, T. Behne, and H. Moll, "In search of the uniquely human", *Behavioral and Brain Sciences*, vol. 28, no. 5, pp. 721–735, Oct. 2005. DOI: 10.1017/S0140525X05540123.

[85] J. O. Bailey, J. N. Bailenson, J. Obradović, and N. R. Aguiar, "Virtual reality's effect on children's inhibitory control, social compliance, and sharing", *Journal of Applied Developmental Psychology*, vol. 64, p. 101 052, Jul. 2019. DOI: 10.1016/j.appdev.2019.101052.

[86] P. Mitchell, S. Parsons, and A. Leonard, "Using virtual environments for teaching social understanding to 6 adolescents with autistic spectrum disorders", *Journal of Autism and Developmental Disorders*, vol. 37, no. 3, pp. 589–600, Mar. 1, 2007. DOI: 10.1007/s10803-006-0189-8.

[87] J. Blascovich, "Social influence within immersive virtual environments", in *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, R. Schroeder, Ed., London: Springer, 2002, pp. 127–145. DOI: 10.1007/978-1-4471-0277-9_8.

[88] P. Huguet, M. P. Galvaing, J. M. Monteil, and F. Dumas, "Social presence effects in the stroop task: Further evidence for an attentional view of social facilitation", *Journal of Personality and Social Psychology*, vol. 77, no. 5, pp. 1011–1025, Nov. 1999. DOI: 10.1037//0022-3514.77.5.1011.

[89] W. Jarrold, P. Mundy, M. Gwaltney, J. Bailenson, N. Hatt, N. McIntyre, K. Kim, M. Solomon, S. Novotny, and L. Swain, "Social attention in a virtual public speaking task in higher functioning children with autism", *Autism Research*, vol. 6, no. 5, pp. 393–410, 2013. DOI: 10.1002/aur.1302.

[90] B. Hamre and R. C. Pianta, "Classroom environments and developmental processes: Conceptualization and measurement", *Handbook of Research on Schools, Schooling and Human Development*, pp. 25–41, 2010.

[91] J. N. Bailenson, N. Yee, J. Blascovich, A. C. Beall, N. Lundblad, and M. Jin, "The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context", *Journal of the Learning Sciences*, vol. 17, no. 1, pp. 102–141, 2008. DOI: 10.1080/10508400701793141.

[92] S. Bioulac, S. Lallemand, A. Rizzo, P. Philip, C. Fabrigoule, and M. Bouvard, "Impact of time on task on ADHD patient's performances in a virtual classroom.", *European Journal of Paediatric Neurology*, vol. 16 5, pp. 514–21, 2012. DOI: https://doi.org/10.1016/j.ejpn.2012.01.006.

[93] U. Díaz-Orueta, C. Garcia-López, N. Crespo-Eguílaz, R. Sánchez-Carpintero, G. Climent, and J. Narbona, "AULA virtual reality test as an attention measure: Convergent validity with conners' continuous performance test", *Child Neuropsychology*, vol. 20, no. 3, pp. 328–342, 2014. DOI: 10.1080/09297049.2013.792332.

[94] A. Rizzo, T. Bowerly, J. Buckwalter, D. Klimchuk, R. Mitura, and T. Parsons, "A virtual reality scenario for all seasons: The virtual classroom", *CNS spectrums*, vol. 11, no. 1, pp. 35–44, 2005. DOI: 10.1017/S1092852900024196.

[95] R. Adams, P. Finn, E. Moes, K. Flannery, and A. " Rizzo, "Distractibility in attention deficit hyperactivity disorder (ADHD): The virtual reality classroom", *Child Neuropsychology*, vol. 15, no. 2, pp. 120–135, 2009. DOI: 10.1080/09297040802169077.

[96] P. Nolin, A. Stipanicic, M. Henry, Y. Lachapelle, D. Lussier-Desrochers, A. Rizzo, and P. Allain, "ClinicaVR: Classroom-CPT: A virtual reality tool for assessing attention

and inhibition in children and adolescents", *Computers in Human Behavior*, vol. 59, pp. 327–333, 2016. DOI: 10.1016/j.chb.2016.02.023.

[97] A. Mangalmurti, W. D. Kistler, B. Quarrie, W. Sharp, S. Persky, and P. Shaw, "Using virtual reality to define the mechanisms linking symptoms with cognitive deficits in attention deficit hyperactivity disorder", *Scientific Reports*, vol. 10, no. 529, 2020. DOI: 10.1038/s41598-019-56936-4.

[98] C. Llinares Millán, J. L. Higuera-Trujillo, A. Montañana i Aviñó, J. Torres, and C. Sentieri, "The influence of classroom width on attention and memory: Virtual-reality-based task performance and neurophysiological effects", *Building Research & Information*, vol. 49, no. 7, pp. 813–826, Oct. 3, 2021. DOI: 10.1080/09613218.2021.1899798.

[99] W. Li, X. Ren, L. Qian, H. Luo, and B. Liu, "Uncovering the effect of classroom climates on learning experience and performance in a virtual environment", *Interactive Learning Environments*, pp. 1–14, Apr. 1, 2023. DOI: 10.1080/10494820.2023.2195450.

[100] R. J. Matheis, M. T. Schultheis, L. A. Tiersky, J. DeLuca, S. R. Millis, and A. Rizzo, "Is learning and memory different in a virtual environment?", *The Clinical Neuropsychologist*, vol. 21, no. 1, pp. 146–161, 2007. DOI: 10.1080/13854040601100668.

[101] G. Makransky, N. K. Andreasen, S. Baceviciute, and R. Mayer, "Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality", *Journal of Educational Psychology*, vol. 113, no. 4, pp. 719–735, Feb. 1, 2021. DOI: 10.1037/edu0000473.

[102] G. Makransky, T. S. Terkildsen, and R. E. Mayer, "Adding immersive virtual reality to a science lab simulation causes more presence but less learning", *Learning and Instruction*, vol. 60, pp. 225–236, Apr. 1, 2019. DOI: 10.1016/j.learninstruc.2017.12.007.

[103] R. Lachman, J. L. Lachman, and E. C. Butterfield, *Cognitive Psychology and Information Processing: An Introduction*. New York: Psychology Press, Jan. 4, 2016, 592 pp. DOI: 10.4324/9781315798844.

[104] M. S. Gazzaniga, *The Cognitive Neurosciences*. MIT Press, 2004, 1480 pp.

[105] C.-C. Wu, F. Wick, and M. Pomplun, "Guidance of visual attention by semantic information in real-world scenes", *Frontiers in Psychology*, vol. 5, 2014. DOI: https://doi.org/10.3389/fpsyg.2014.00054.

[106] F. Hutmacher, "Why is there so much more research on vision than on any other sensory modality?", *Frontiers in Psychology*, vol. 10, p. 2246, 2019. DOI: 10.3389/fpsyg.2019.02246.

[107] A. M. Treisman and G. Gelade, "A feature-integration theory of attention", *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, Jan. 1, 1980. DOI: 10.1016/0010-0285(80)90005-5.

[108] E. Awh, E. K. Vogel, and S. Oh, "Interactions between attention and working memory", *Neuroscience*, vol. 139, no. 1, pp. 201–208, 2006. DOI: https://doi.org/10.1016/j.neuroscience.2005.08.023.

[109] C. N. L. Olivers and P. R. Roelfsema, "Attention for action in visual working memory", *Cortex*, vol. 131, pp. 179–194, Oct. 1, 2020. DOI: 10.1016/j.cortex.2020.07.011.

[110] M. Carrasco, "Visual attention: The past 25 years", *Vision Research*, vol. 51, no. 13, pp. 1484–1525, 2011. DOI: https://doi.org/10.1016/j.visres.2011.04.012.

[111] D. M. Beck and S. Kastner, "Top-down and bottom-up mechanisms in biasing competition in the human brain", *Vision Research*, vol. 49, no. 10, pp. 1154–1165, 2009. DOI: https://doi.org/10.1016/j.visres.2008.07.012.

[112] S. Hutt, C. Mills, N. Bosch, K. Krasich, J. Brockmole, and S. D'Mello, ""out of the fr-eyeing pan": Towards gaze-based models of attention during learning with technology in the classroom", in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, ser. UMAP '17, New York, NY, USA: ACM, 2017, pp. 94–103. DOI: 10.1145/3079628.3079669.

[113] R. J. Jacobs, "Visual resolution and contour interaction in the fovea and periphery", *Vision Research*, vol. 19, no. 11, pp. 1187–1195, 1979. DOI: https://doi.org/10.1016/0042-6989(79)90183-4.

[114] T. C. Kübler, C. Rothe, U. Schiefer, W. Rosenstiel, and E. Kasneci, "SubsMatch 2.0: Scanpath comparison and classification based on subsequence frequencies", *Behavior Research Methods*, vol. 49, no. 3, pp. 1048–1064, 2017. DOI: https://doi.org/10.3758/s13428-016-0765-6.

[115] K. Holmqvist and R. Andersson, *Eye Tracking: A Comprehensive Guide to Methods, Paradigms, and Measures*. Lund, Sweden: Lund Eye-Tracking Research Institute, 2017, 746 pp.

[116] M. A. Just and P. A. Carpenter, "A theory of reading: From eye fixations to comprehension", *Psychological Review*, vol. 87, pp. 329–354, 1980. DOI: 10.1037/0033-295X.87.4.329.

[117] H. Deubel, "The time course of presaccadic attention shifts", *Psychological Research*, vol. 72, no. 6, pp. 630–640, Nov. 1, 2008. DOI: 10.1007/s00426-008-0165-3.

[118] C. E. Connor, H. E. Egeth, and S. Yantis, "Visual attention: Bottom-up versus top-down", *Current Biology*, vol. 14, no. 19, pp. 850–852, 2004. DOI: https://doi.org/10.1016/j.cub.2004.09.041.

[119] A. L. Yarbus, *Eye Movements and Vision*, 1st ed. New York, NY, US: Springer, 1967, 222 pp.

[120] B. W. Tatler, N. J. Wade, H. Kwan, J. M. Findlay, and B. M. Velichkovsky, "Yarbus, eye movements, and vision", *i-Perception*, vol. 1, no. 1, pp. 7–27, Apr. 1, 2010. DOI: 10.1068/i0382.

[121] S. Ahern and J. Beatty, "Pupillary responses during information processing vary with scholastic aptitude test scores", *Science*, vol. 205, no. 4412, pp. 1289–1292, Sep. 21, 1979. DOI: 10.1126/science.472746.

[122] M. M. Bradley, L. Miccoli, M. A. Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation", *Psychophysiology*, vol. 45, no. 4, pp. 602–607, Jul. 2008. DOI: 10.1111/j.1469-8986.2008.00654.x.

[123] J. L. Bradshaw, "Pupil size and problem solving", *Quarterly Journal of Experimental Psychology*, vol. 20, no. 2, pp. 116–122, May 1, 1968. DOI: 10.1080/14640746808400139.

[124] O. White and R. M. French, "Pupil diameter may reflect motor control and learning", *Journal of Motor Behavior*, vol. 49, no. 2, pp. 141–149, Mar. 4, 2017. DOI: 10.1080/00222895.2016.1161593.

[125] A. Felnhofer, O. D. Kothgassner, M. Schmidt, A.-K. Heinzle, L. Beutl, H. Hlavacs, and I. Kryspin-Exner, "Is virtual reality emotionally arousing? investigating five emotion inducing virtual park scenarios", *International Journal of Human-Computer Studies*, vol. 82, pp. 48–56, Oct. 1, 2015. DOI: 10.1016/j.ijhcs.2015.05.004.

[126] M. Fanourakis and G. Chanel, "Attenuation of the dynamic pupil light response during screen viewing for arousal assessment", *Frontiers in Virtual Reality*, vol. 3, 2022. DOI: 10.3389/frvir.2022.971613.

[127] J. Beatty, "Task-evoked pupillary responses, processing load, and the structure of processing resources", *Psychological Bulletin*, vol. 91, no. 2, pp. 276–292, Mar. 1982. DOI: 10.1037/0033-2909.91.2.276.

[128] D. Kahneman and J. Beatty, "Pupil diameter and load on memory", *Science*, vol. 154, no. 3756, pp. 1583–1585, Dec. 23, 1966. DOI: 10.1126/science.154.3756.1583.

[129] K. Rayner, "Eye movements and attention in reading, scene perception, and visual search", *Quarterly Journal of Experimental Psychology (2006)*, vol. 62, no. 8, pp. 1457–1506, Aug. 2009. DOI: 10.1080/17470210902816461.

[130] M.-J. Tsai, H.-T. Hou, M.-L. Lai, W.-Y. Liu, and F.-Y. Yang, "Visual attention for solving multiple-choice science problem: An eye-tracking analysis", *Computers & Education*, vol. 58, no. 1, pp. 375–385, Jan. 1, 2012. DOI: 10.1016/j.compedu.2011.07.012.

[131] M. A. Just and P. A. Carpenter, "Eye fixations and cognitive processes", *Cognitive Psychology*, vol. 8, no. 4, pp. 441–480, Oct. 1, 1976. DOI: 10.1016/0010-0285(76)90015-3.

[132] M. A. Just and P. A. Carpenter, "Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability", *Psychological Review*, vol. 92, pp. 137–172, Apr. 1985. DOI: 10.1037/0033-295X.92.2.137.

[133] S. Zhu, K. J. Lakshminarasimhan, N. Arfaei, and D. E. Angelaki, "Eye movements reveal spatiotemporal dynamics of visually-informed planning in navigation", *eLife*, vol. 11, e73097, May 3, 2022. DOI: 10.7554/eLife.73097.

[134] K. Nakayama and P. Martini, "Situating visual search", *Vision Research*, vol. 51, no. 13, pp. 1526–1537, 2011. DOI: https://doi.org/10.1016/j.visres.2010.09.003.

[135] M. L. Mele and S. Federici, "Gaze and eye-tracking solutions for psychological research", *Cognitive Processing*, vol. 13, no. 1, pp. 261–265, Aug. 1, 2012. DOI: 10.1007/s10339-012-0499-z.

[136] X. Ma, Y. Liu, R. Clariana, C. Gu, and P. Li, "From eye movements to scanpath networks: A method for studying individual differences in expository text reading", *Behavior Research Methods*, vol. 55, no. 2, pp. 730–750, Feb. 1, 2023. DOI: 10.3758/s13428-022-01842-3.

[137] S. Stranc and K. Muldner, "Scanpath analysis of student attention during problem solving with worked examples", *Artificial Intelligence in Education*, vol. 12164, pp. 306–311, Jun. 10, 2020. DOI: 10.1007/978-3-030-52240-7_56.

[138] G. E. Raptis, C. A. Fidas, and N. M. Avouris, "On implicit elicitation of cognitive strategies using gaze transition entropies in pattern recognition tasks", in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '17, New York, NY, USA: ACM, 2017, pp. 1993–2000. DOI: 10.1145/3027063.3053106.

[139] C. Kosel, D. Holzberger, and T. Seidel, "Identifying expert and novice visual scanpath patterns and their relationship to assessing learning-relevant student characteristics", *Frontiers in Education*, vol. 5, 2021. DOI: 10.3389/feduc.2020.612175.

[140] N. A. McIntyre and T. Foulsham, "Scanpath analysis of expertise and culture in teacher gaze in real-world classrooms", *Instructional Science*, vol. 46, no. 3, pp. 435–455, Jun. 1, 2018. DOI: 10.1007/s11251-017-9445-x.

[141] L. L. Di Stasi, C. Diaz-Piedra, H. Rieiro, J. M. Sánchez Carrión, M. Martin Berrido, G. Olivares, and A. Catena, "Gaze entropy reflects surgical task load", *Surgical Endoscopy*, vol. 30, no. 11, pp. 5034–5043, Nov. 1, 2016. DOI: 10.1007/s00464-016-4851-8.

[142] S. Mejia-Romero, J. Michaels, J. Eduardo Lugo, D. Bernardin, and J. Faubert, "Gaze movement's entropy analysis to detect workload levels", in *Proceedings of International Conference on Trends in Computational and Cognitive Engineering*, M. S. Kaiser, A. Bandyopadhyay, M. Mahmud, and K. Ray, Eds., ser. Advances in Intelligent Systems and Computing, vol. 1309, Singapore: Springer, 2021, pp. 147–154. DOI: 10.1007/978-981-33-4673-4_13.

[143] B. Shiferaw, L. Downey, and D. Crewther, "A review of gaze entropy as a measure of visual scanning efficiency", *Neuroscience & Biobehavioral Reviews*, vol. 96, pp. 353–366, Jan. 1, 2019. DOI: 10.1016/j.neubiorev.2018.12.007.

[144] A. Frischen, A. P. Bayliss, and S. P. Tipper, "Gaze cueing of attention: Visual attention, social cognition, and individual differences", *Psychological Bulletin*, vol. 133, no. 4, pp. 694–724, 2007. DOI: 10.1037/0033-2909.133.4.694.

[145] P. Yazdan-Shahmorad, N. Sammaknejad, and F. Bakouie, "Graph-based analysis of visual scanning patterns: A developmental study on green and normal images", *Scientific Reports*, vol. 10, no. 1, p. 7791, 2020. DOI: 10.1038/s41598-020-63951-3.

[146] M. Sadria, S. Karimi, and A. T. Layton, "Network centrality analysis of eye-gaze data in autism spectrum disorder", *Computers in Biology and Medicine*, vol. 111, p. 103 332, Aug. 1, 2019. DOI: https://doi.org/10.1016/j.compbiomed.2019.103332.

[147] Q. Guillon, M. H. Afzali, B. Rogé, S. Baduel, J. Kruck, and N. Hadjikhani, "The importance of networking in autism gaze analysis", *PLOS ONE*, vol. 10, no. 10, e0141191, Oct. 23, 2015. DOI: 10.1371/journal.pone.0141191.

[148] S. Andrist, W. Collier, M. Gleicher, B. Mutlu, and D. Shaffer, "Look together: Analyzing gaze coordination with epistemic network analysis", *Frontiers in Psychology*, vol. 6, 2015. DOI: https://doi.org/10.3389/fpsyg.2015.01016.

[149] S. Andrist, A. Ruis, and D. Shaffer, "A network analytic approach to gaze coordination during a collaborative task", *Computers in Human Behavior*, vol. 89, Jul. 1, 2018. DOI: 10.1016/j.chb.2018.07.017.

[150] B. Schneider, S. Abu-El-Haija, J. Reesman, and R. Pea, "Toward collaboration sensing: Applying network analysis techniques to collaborative eye-tracking data", in *ACM International Conference Proceeding Series*, 2013, pp. 107–111. DOI: 10.1145/2460296.2460317.

[151] R.-M. Rahal and S. Fiedler, "Understanding cognitive and affective mechanisms in social psychology through eye-tracking", *Journal of Experimental Social Psychology*, vol. 85, p. 103 842, Nov. 1, 2019. DOI: 10.1016/j.jesp.2019.103842.

[152] R. Jacob and K. Karn, "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises", in *Mind; a Quarterly Review of Psychology and Philosophy*, vol. 2, Jan. 1, 2003, pp. 573–605. DOI: 10.1016/B978-044451020-4/50031-1.

[153] L. Hahn and P. Klein, "Eye tracking in physics education research: A systematic literature review", *Physical Review Physics Education Research*, vol. 18, no. 1, p. 013 102, Mar. 2, 2022. DOI: 10.1103/PhysRevPhysEducRes.18.013102.

[154] A. Rouinfar, E. Agra, A. M. Larson, N. S. Rebello, and L. C. Loschky, "Linking attentional processes and conceptual problem solving: Visual cues facilitate the automaticity of extracting relevant information from diagrams", *Frontiers in Psychology*, vol. 5, p. 1094, 2014. DOI: https://doi.org/10.3389/fpsyg.2014.01094.

[155] K. Sharma, M. Giannakos, and P. Dillenbourg, "Eye-tracking and artificial intelligence to enhance motivation and learning", *Smart Learning Environments*, vol. 7, no. 1, p. 13, Apr. 26, 2020. DOI: 10.1186/s40561-020-00122-x.

[156] H. Jarodzka, I. Skuballa, and H. Gruber, "Eye-tracking in educational practice: Investigating visual perception underlying teaching and learning in the classroom", *Educational Psychology Review*, vol. 33, no. 1, pp. 1–10, Mar. 1, 2021. DOI: 10.1007/s10648-020-09565-7.

[157] K. Krejtz, A. T. Duchowski, I. Krejtz, A. Kopacz, and P. Chrząstowski-Wachtel, "Gaze transitions when learning with multimedia", *Journal of Eye Movement Research*, vol. 9, no. 1, Feb. 10, 2016. DOI: 10.16910/jemr.9.1.5.

[158] M. Tomasello, "Joint attention as social cognition", in *Joint Attention: Its Origins and Role in Development*, C. Moore and P. J. Dunham, Eds., Lawrence Erlbaum Associates, Inc, 1995, pp. 103–130.

[159] D. Lundqvist and A. Ohman, "Emotion regulates attention: The relation between facial configurations, facial emotion, and visual attention", *Visual Cognition*, vol. 12, no. 1, pp. 51–84, 2005. DOI: 10.1080/13506280444000085.

[160] J. L. Rosch and J. J. Vogel-Walcutt, "A review of eye-tracking applications as tools for training", *Cognition, Technology & Work*, vol. 15, no. 3, pp. 313–327, Aug. 1, 2013. DOI: 10.1007/s10111-012-0234-7.

[161] X. Huang, Q. Zhao, Y. Liu, D. Harris, and M. Shawler, "Learning in an immersive VR environment: Role of learner characteristics and relations between learning and psychological outcomes", *Journal of Educational Technology Systems*, Dec. 3, 2023. DOI: 10.1177/00472395231216943.

[162] C. Song, S.-Y. Shin, and K.-S. Shin, "Optimizing foreign language learning in virtual reality: A comprehensive theoretical framework based on constructivism and cognitive load theory (VR-CCL)", *Appl. Sci*, vol. 13, no. 23, p. 12 557, Sep. 29, 2023. DOI: 10.3390/app132312557.

[163] W.-S. Wang, C.-J. Lin, H.-Y. Lee, T.-T. Wu, and Y.-M. Huang, "Feedback mechanism in immersive virtual reality influences physical hands-on task performance and cognitive load", *International Journal of Human–Computer Interaction*, pp. 1–13, May 11, 2023. DOI: 10.1080/10447318.2023.2209837.

[164] P. Peng, T. Wang, C. Wang, and X. Lin, "A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status", *Psychological Bulletin*, vol. 145, pp. 189–236, 2019. DOI: 10.1037/bul0000182.

[165] H.-H. Choi, J. J. G. van Merriënboer, and F. Paas, "Effects of the physical environment on cognitive load and learning: Towards a new model of cognitive load", *Educational Psychology Review*, vol. 26, no. 2, pp. 225–244, Jun. 1, 2014. DOI: 10.1007/s10648-014-9262-6.

[166] P. A. Kirschner, "Cognitive load theory: Implications of cognitive load theory on the design of learning", *Learning and Instruction*, vol. 12, no. 1, pp. 1–10, Feb. 1, 2002. DOI: 10.1016/S0959-4752(01)00014-7.

[167] J.-C. Woo, "Digital game-based learning supports student motivation, cognitive success, and performance outcomes", *Journal of Educational Technology & Society*, vol. 17, no. 3, pp. 291–307, 2014. DOI: https://www.jstor.org/stable/jeductechsoci.17.3.291.

[168] "VIVE pro eye user guide", HTC, manual, 2020.

[169] A. Sipatchin, S. Wahl, and K. Rifai, "Accuracy and precision of the HTC VIVE PRO eye tracking in head-restrained and head-free conditions", *Investigative Ophthalmology & Visual Science*, vol. 61, no. 7, p. 5071, Jun. 2020.

[170] A. Sipatchin, M. García García, and S. Wahl, "Impact of unconstrained head movements to scotoma and enhanced scotoma simulation in virtual-reality (VR) smooth pursuit gaming", *Investigative Ophthalmology & Visual Science*, vol. 62, no. 8, pp. 1446–1446, Jun. 21, 2021.

[171] "EyeLink 1000 user manual", EyeLink, Mississauga, Ontario, Canada, Version 1.5.2, 2005.

[172] M. Juhola, V. Jäntti, and I. Pyykkö, "Effect of sampling frequencies on computation of the maximum velocity of saccadic eye movements", *Biological Cybernetics*, vol. 53, no. 2, pp. 67–72, Dec. 1, 1985. DOI: 10.1007/BF00337023.

[173] A. D. Souchet, S. Philippe, D. Lourdeaux, and L. Leroy, "Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review", *International Journal of Human–Computer Interaction*, vol. 38, no. 9, pp. 801–824, May 28, 2022. DOI: 10.1080/10447318.2021.1976509.

[174] J. Moreno-Arjonilla, A. López-Ruiz, J. R. Jiménez-Pérez, J. E. Callejas-Aguilera, and J. M. Jurado, "Eye-tracking on virtual reality: A survey", *Virtual Reality*, vol. 28, no. 1, p. 38, Feb. 5, 2024. DOI: 10.1007/s10055-023-00903-y.

[175] R. Soret, P. Charras, I. Khazar, C. Hurter, and V. Peysakhovich, "Eye-tracking and virtual reality in 360-degrees: Exploring two ways to assess attentional orienting in rear space", in *ACM Symposium on Eye Tracking Research and Applications*, ser. ETRA '20 Adjunct, New York, NY, USA: ACM, 2020. DOI: 10.1145/3379157.3391418.

[176] L. Rokach, "Ensemble-based classifiers", *Artificial Intelligence Review*, vol. 33, no. 1, pp. 1–39, Feb. 1, 2010. DOI: 10.1007/s10462-009-9124-7.

[177] E. Kasneci, G. Kasneci, U. Trautwein, T. Appel, M. Tibus, S. M. Jaeggi, and P. Gerjets, "Do your eye movements reveal your performance on an IQ test? a study linking eye movements and socio-demographic information to fluid intelligence", *PLOS ONE*, vol. 17, no. 3, e0264316, Mar. 2022. DOI: 10.1371/journal.pone.0264316.

[178] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable AI for trees", *Nature Machine Intelligence*, vol. 2, no. 1, pp. 56–67, Jan. 2020. DOI: 10.1038/s42256-019-0138-9.

[179] N. Alghamdi and W. Alhalabi, "Fixation detection with ray-casting in immersive virtual reality", *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 7, 2019. DOI: 10.14569/IJACSA.2019.0100710.

[180] A. T. Duchowski, "Gaze-based interaction: A 30 year retrospective", *Computers & Graphics*, vol. 73, pp. 59–69, Jun. 1, 2018. DOI: 10.1016/j.cag.2018.04.002.

[181] K. Krejtz, A. Duchowski, T. Szmidt, I. Krejtz, F. González Perilli, A. Pires, A. Vilaro, and N. Villalobos, "Gaze transition entropy", *ACM Transactions on Applied Perception*, vol. 13, no. 1, pp. 1–20, 2015. DOI: 10.1145/2834121.

[182] R. Diestel, *Graph Theory* (Graduate Texts in Mathematics), 5th ed. Berlin Heidelberg: Springer-Verlag, 2017.

[183] L. Candeloro, L. Savini, and A. Conte, "A new weighted degree centrality measure: The application in an animal disease epidemic", *PLOS ONE*, vol. 11, no. 11, e0165781, Nov. 1, 2016. DOI: 10.1371/journal.pone.0165781.

[184] S. P. Borgatti, "Centrality and network flow", *Social Networks*, vol. 27, no. 1, pp. 55–71, Jan. 1, 2005. DOI: 10.1016/j.socnet.2004.11.008.

[185] M. Tantardini, F. Ieva, L. Tajoli, and C. Piccardi, "Comparing methods for comparing networks", *Scientific Reports*, vol. 9, no. 1, p. 17 557, Nov. 26, 2019. DOI: 10.1038/s41598-019-53708-y.

[186] C. S. Q. Siew, D. U. Wulff, N. M. Beckage, and Y. N. Kenett, "Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics", *Complexity*, vol. 2019, e2108423, Jun. 17, 2019. DOI: 10.1155/2019/2108423.

[187] O. Sporns, "Graph theory methods: Applications in brain networks", *Dialogues in Clinical Neuroscience*, vol. 20, no. 2, pp. 111–121, Jun. 30, 2018. DOI: 10.31887/DCNS.2018.20.2/osporns.

[188] M. Zhu and G. Feng, "An exploratory study using social network analysis to model eye movements in mathematics problem solving", in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, ser. LAK '15, New York, NY, USA: ACM, 2015, pp. 383–387. DOI: 10.1145/2723576.2723591.

[189] J. Xue, C. Li, C. Quan, Y. Lu, J. Yue, and C. Zhang, "Uncovering the cognitive processes underlying mental rotation: An eye-movement study", *Scientific Reports*, vol. 7, no. 1, p. 10 076, Aug. 30, 2017. DOI: 10.1038/s41598-017-10683-6.

[190] E. H. Hess and J. M. Polt, "Pupil size in relation to mental activity during simple problem-solving", *Science*, vol. 143, no. 3611, pp. 1190–1192, Mar. 13, 1964. DOI: 10.1126/science.143.3611.1190.

[191] M. A. Just and P. A. Carpenter, "The intensity dimension of thought: Pupillometric indices of sentence processing", *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale*, vol. 47, pp. 310–339, 1993. DOI: 10.1037/h0078820.

[192] K. Krejtz, A. T. Duchowski, A. Niedzielska, C. Biele, and I. Krejtz, "Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze", *PLOS ONE*, vol. 13, no. 9, e0203629, Sep. 14, 2018. DOI: 10.1371/journal.pone.0203629.

[193] N. Unsworth and M. K. Robison, "Tracking arousal state and mind wandering with pupillometry", *Cognitive, Affective, & Behavioral Neuroscience*, vol. 18, no. 4, pp. 638–664, Aug. 1, 2018. DOI: 10.3758/s13415-018-0594-4.

[194] R. Hershman, D. Milshtein, and A. Henik, "The contribution of temporal analysis of pupillometry measurements to cognitive research", *Psychological Research*, vol. 87, no. 1, pp. 28–42, Feb. 2023. DOI: 10.1007/s00426-022-01656-0.

[195] K. Pietroszek, "Raycasting in virtual reality", in *Encyclopedia of Computer Graphics and Games*, N. Lee, Ed., Cham: Springer International Publishing, 2018, pp. 1–3.

[196] C. R. Piontkowski D., "Attention in the classroom", in *Attention and Cognitive Development*, L. M. Hale G.A., Ed., Springer, Boston, MA, 1979. DOI: https://doi.org/10.1007/978-1-4613-2985-511.

[197] K. Krejtz, T. Szmidt, A. T. Duchowski, and I. Krejtz, "Entropy-based statistical analysis of eye movement transitions", in *Proceedings of the Symposium on Eye Tracking Research and Applications*, ser. ETRA '14, New York, NY, USA: ACM, 2014, pp. 159–166. DOI: 10.1145/2578153.2578176.

[198] A. Coutrot, J. H. Hsiao, and A. B. Chan, "Scanpath modeling and classification with hidden markov models", *Behavior Research Methods*, vol. 50, no. 1, pp. 362–379, 2018. DOI: https://doi.org/10.3758/s13428-017-0876-8.

[199] F. Cristino, S. Mathôt, J. Theeuwes, and I. D. Gilchrist, "ScanMatch: A novel method for comparing fixation sequences", *Behavior Research Methods*, vol. 42, no. 3, pp. 692–700, 2010. DOI: https://doi.org/10.3758/BRM.42.3.692.

[200] U. Brandes and T. Erlebach, Eds., *Network Analysis*, vol. 3418, Lecture Notes in Computer Science, Berlin, Heidelberg: Springer, 2005, 472 pp. DOI: 10.1007/b106453.

[201] K. M. Curtin, "Network analysis", in *Comprehensive Geographic Information Systems*, B. Huang, Ed., Oxford: Elsevier, Jan. 1, 2018, pp. 153–161.

[202] R. Göllner, R. I. Damian, B. Nagengast, B. W. Roberts, and U. Trautwein, "It's not only who you are but who you are with: High school composition and individuals' attainment over the life course", *Psychological Science*, vol. 29, no. 11, pp. 1785–1796, 2018. DOI: 10.1177/0956797618794454.

[203] A. C. Frenzel, R. Pekrun, and T. Goetz, "Perceived learning environment and students' emotional experiences: A multilevel analysis of mathematics classrooms", *Learning and Instruction*, vol. 17, no. 5, pp. 478–493, Oct. 1, 2007. DOI: 10.1016/j.learninstruc.2007.09.001.

[204] F. Heider and M. Simmel, "An experimental study of apparent behavior", *The American Journal of Psychology*, vol. 57, no. 2, pp. 243–259, 1944. DOI: 10.2307/1416950.

[205] C. Stein, "Uncanny valley in virtual reality", in *Encyclopedia of Computer Graphics and Games*, N. Lee, Ed., Cham: Springer International Publishing, 2018, pp. 1–3.

[206] L. Rebenitsch and C. Owen, "Review on cybersickness in applications and visual displays", *Virtual Reality*, vol. 20, no. 2, pp. 101–125, 2016. DOI: https://doi.org/10.1007/s10055-016-0285-9.

[207] D. Voyer, P. Jansen, and S. Kaltner, "Mental rotation with egocentric and object-based transformations", *Quarterly Journal of Experimental Psychology*, vol. 70, no. 11, pp. 2319–2330, Nov. 2017. DOI: 10.1080/17470218.2016.1233571.

[208] M. Wraga, W. L. Thompson, N. M. Alpert, and S. M. Kosslyn, "Implicit transfer of motor strategies in mental rotation", *Brain and Cognition*, vol. 52, no. 2, pp. 135–143, 2003. DOI: 10.1016/S0278-2626(03)00033-2.

[209] A. L. Gardony, H. A. Taylor, and T. T. Brunyé, "What does physical rotation reveal about mental rotation?", *Psychological Science*, vol. 25, no. 2, pp. 605–612, 2014. DOI: 10.1177/0956797613503174.

[210] B. Chavez and S. Bayona, "Virtual reality in the learning process", in *Trends and Advances in Information Systems and Technologies*, Á. Rocha, H. Adeli, L. P. Reis, and S. Costanzo, Eds., ser. Advances in Intelligent Systems and Computing, Cham: Springer International Publishing, 2018, pp. 1345–1356. DOI: 10.1007/978-3-319-77712-2_129.

[211] "Dispositif(apparatus)", in *The Cambridge Foucault Lexicon*, J. Nale and L. Lawlor, Eds., Cambridge: Cambridge University Press, 2014, pp. 126–132.

[212] S. M. Tapley and M. P. Bryden, "An investigation of sex differences in spatial ability: Mental rotation of three-dimensional objects", *Canadian Journal of Psychology/Revue canadienne de psychologie*, vol. 31, pp. 122–130, 1977. DOI: 10.1037/h0081655.

[213] S. G. Vandenberg and A. R. Kuse, "Mental rotations, a group test of three-dimensional spatial visualization", *Perceptual and Motor Skills*, vol. 47, no. 2, pp. 599–604, 1978. DOI: 10.2466/pms.1978.47.2.599.

[214] M. C. Linn and A. C. Petersen, "Emergence and characterization of sex differences in spatial ability: A meta-analysis", *Child Development*, vol. 56, no. 6, pp. 1479–1498, 1985. DOI: 10.2307/1130467.

[215] M. E. Desrocher, M. L. Smith, and M. J. Taylor, "Stimulus and sex-differences in performance of mental rotation - evidence from event-related potentials", *Brain and Cognition*, vol. 28, no. 1, pp. 14–38, Jun. 1, 1995. DOI: 10.1006/brcg.1995.1031.

[216] T. D. Parsons, P. Larson, K. Kratz, M. Thiebaux, B. Bluestein, J. G. Buckwalter, and A. A. Rizzo, "Sex differences in mental rotation and spatial rotation in a virtual environment", *Neuropsychologia*, vol. 42, no. 4, pp. 555–562, 2004. DOI: 10.1016/j.neuropsychologia.2003.08.014.

[217] I. Lochhead, N. Hedley, A. Çöltekin, and B. Fisher, "The immersive mental rotations test: Evaluating spatial ability in virtual reality", *Frontiers in Virtual Reality*, vol. 3, 2022. DOI: 10.3389/frvir.2022.820237.

[218] W. Fuhl, E. Bozkir, and E. Kasneci, "Reinforcement learning for the privacy preservation and manipulation of eye tracking data", in *Artificial Neural Networks and Machine Learning – ICANN 2021*, I. Farkaš, P. Masulli, S. Otte, and S. Wermter, Eds., Cham: Springer International Publishing, 2021, pp. 595–607. DOI: 10.1007/978-3-030-86380-7_48.

[219] E. Bozkir, O. Günlü, W. Fuhl, R. F. Schaefer, and E. Kasneci, "Differential privacy for eye tracking with temporal correlations", *PLOS ONE*, vol. 16, no. 8, e0255979, Aug. 17, 2021. DOI: 10.1371/journal.pone.0255979.

[220] R. Marshall, A. Pardo, D. Smith, and T. Watson, "Implementing next generation privacy and ethics research in education technology", *British Journal of Educational Technology*, vol. 53, no. 4, pp. 737–755, Jul. 2022. DOI: 10.1111/bjet.13224.

[221] E. Bozkir, A. B. Ünal, M. Akgün, E. Kasneci, and N. Pfeifer, "Privacy preserving gaze estimation using synthetic images via a randomized encoding based framework", in *ACM Symposium on Eye Tracking Research and Applications*, ser. ETRA '20 Short Papers, New York, NY, USA: ACM, Jun. 2, 2020, pp. 1–5. DOI: 10.1145/3379156.3391364.

[222] M. Wang, A. Bodonhelyi, E. Bozkir, and E. Kasneci, "TurboSVM-FL: Boosting federated learning through SVM aggregation for lazy clients", in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, Mar. 24, 2024, pp. 15 546–15 554. DOI: 10.1609/aaai.v38i14.29481.

[223] C. Li, C. Zhang, A. Waghwase, L.-H. Lee, F. Rameau, Y. Yang, S.-H. Bae, and C. S. Hong, *Generative AI meets 3d: A survey on text-to-3d in AIGC era*, May 26, 2023. DOI: 10.48550/arXiv.2305.06131.

[224] G. Ganis and R. Kievit, "A new set of three-dimensional shapes for investigating mental rotation processes: Validation data and stimulus set", *Journal of Open Psychology Data*, vol. 3, 2015. DOI: 10.5334/jopd.ai.

[225] N. Hoyek, C. Collet, P. Fargier, and A. Guillot, "The use of the vandenberg and kuse mental rotation test in children", *Journal of Individual Differences*, vol. 33, no. 1, pp. 62–67, Jan. 2012. DOI: 10.1027/1614-0001/a000063.

[226] K. C. Moen, M. R. Beck, S. M. Saltzmann, T. M. Cowan, L. M. Burleigh, L. G. Butler, J. Ramanujam, A. S. Cohen, and S. G. Greening, "Strengthening spatial reasoning: Elucidating the attentional and neural mechanisms associated with mental rotation skill development", *Cognitive Research: Principles and Implications*, vol. 5, no. 1, p. 20, 2020. DOI: 10.1186/s41235-020-00211-y.

[227] V. Varriale, M. W. v. d. Molen, and V. D. Pascalis, "Mental rotation and fluid intelligence: A brain potential analysis", *Intelligence*, vol. 69, pp. 146–157, 2018. DOI: 10.1016/j.intell.2018.05.007.

[228] C. Bruce and Z. Hawes, "The role of 2d and 3d mental rotation in mathematics for young children: What is it? why does it matter? and what can we do about it?", *ZDM Mathematics Education*, vol. 47, pp. 331–343, Jun. 2015. DOI: 10.1007/s11858-014-0637-4.

[229] Z. Hawes, J. Moss, B. Caswell, and D. Poliszczuk, "Effects of mental rotation training on children's spatial and mathematics performance: A randomized controlled study", *Trends in Neuroscience and Education*, vol. 4, no. 3, pp. 60–68, 2015. DOI: 10.1016/j.tine.2015.05.001.

[230] R. N. Shepard and J. Metzler, "Mental rotation of three-dimensional objects", *Science*, vol. 171, no. 3972, pp. 701–703, 1971. DOI: 10.1126/science.171.3972.701.

[231] B. B. Van Acker, K. Bombeke, W. Durnez, D. D. Parmentier, J. C. Mateus, A. Biondi, J. Saldien, and P. Vlerick, "Mobile pupillometry in manual assembly: A pilot study exploring the wearability and external validity of a renowned mental workload lab measure", *International Journal of Industrial Ergonomics*, vol. 75, p. 102 891, 2020. DOI: 10.1016/j.ergon.2019.102891.

[232] D. Voyer, "Time limits and gender differences on paper-and-pencil tests of mental rotation: A meta-analysis", *Psychonomic Bulletin & Review*, vol. 18, no. 2, pp. 267–277, 2011. DOI: 10.3758/s13423-010-0042-0.

[233] J. M. Zacks, "Neuroimaging studies of mental rotation: A meta-analysis and review", *Journal of Cognitive Neuroscience*, vol. 20, no. 1, pp. 1–19, 2008. DOI: 10.1162/jocn.2008.20013.

[234] M. Fisher, T. Meredith, and M. Gray, "Sex differences in mental rotation ability are a consequence of procedure and artificiality of stimuli", *Evolutionary Psychological Science*, vol. 4, pp. 1–10, Jun. 2018. DOI: 10.1007/s40806-017-0120-x.

[235] A. J. Toth and M. J. Campbell, "Investigating sex differences, cognitive effort, strategy, and performance on a computerised version of the mental rotations test via eye tracking", *Scientific Reports*, vol. 9, p. 19 430, 2019. DOI: 10.1038/s41598-019-56041-6.

[236] A. R. Bilge and H. A. Taylor, "Framing the figure: Mental rotation revisited in light of cognitive strategies", *Memory & Cognition*, vol. 45, no. 1, pp. 63–80, Jan. 1, 2017. DOI: 10.3758/s13421-016-0648-1.

[237] A. L. Gardony, M. D. Eddy, T. T. Brunyé, and H. A. Taylor, "Cognitive strategies in the mental rotation task revealed by EEG spectral power", *Brain and Cognition*, vol. 118, pp. 1–18, 2017. DOI: 10.1016/j.bandc.2017.07.003.

[238] R. N. Shepard and L. A. Cooper, *Mental images and their transformations*. Cambridge, MA, US: The MIT Press, 1986, viii, 364.

[239] J. E. Lauer, E. Yhang, and S. F. Lourenco, "The development of gender differences in spatial reasoning: A meta-analytic review", *Psychological Bulletin*, vol. 145, no. 6, pp. 537–565, Jun. 2019. DOI: 10.1037/bul0000191.

[240] B. Tomasino and M. Gremese, "Effects of stimulus type and strategy on mental rotation network: An activation likelihood estimation meta-analysis", *Frontiers in Human Neuroscience*, vol. 9, p. 693, 2016. DOI: 10.3389/fnhum.2015.00693.

[241] M. Kozhevnikov and R. Dhond, "Understanding immersivity: Image generation and transformation processes in 3d immersive environments", *Frontiers in Psychology*, vol. 3, 2012. DOI: 10.3389/fpsyg.2012.00284.

[242] G. A. Holleman, I. T. C. Hooge, C. Kemner, and R. S. Hessels, "The 'real-world approach' and its problems: A critique of the term ecological validity", *Frontiers in Psychology*, vol. 11, 2020. DOI: 10.3389/fpsyg.2020.00721.

[243] Y. Aitsiselmi and N. S. Holliman, "Using mental rotation to evaluate the benefits of stereoscopic displays", in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 7237, Feb. 1, 2009. DOI: 10.1117/12.824527.

[244] J. Li, P. Wang, D. Chen, S. Qi, X. Sang, and B. Yan, "Performance evaluation of 3d light field display based on mental rotation tasks", in *VR/AR and 3D Displays*, W. Song and F. Xu, Eds., ser. Communications in Computer and Information Science, Springer, 2021, pp. 33–44. DOI: 10.1007/978-981-33-6549-0_4.

[245] P.-H. Lin and S.-C. Yeh, "How motion-control influences a VR-supported technology for mental rotation learning: From the perspectives of playfulness, gender difference and technology acceptance model", *International Journal of Human-computer Interaction*, vol. 35, no. 18, pp. 1736–1746, Feb. 12, 2019. DOI: 10.1080/10447318.2019.1571784.

[246] L. A. Cooper, "Mental representation of three-dimensional objects in visual problem solving and recognition", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 16, no. 6, pp. 1097–1106, 1990. DOI: 10.1037/0278-7393.16.6.1097.

[247] M. Pittalis and C. Christou, "Coding and decoding representations of 3d shapes", *The Journal of Mathematical Behavior*, vol. 32, no. 3, pp. 673–689, Sep. 1, 2013. DOI: 10.1016/j.jmathb.2013.08.004.

[248] E. B. Goldstein, "Rotation of objects in pictures viewed at an angle: Evidence for different properties of two types of pictorial space", *Journal of experimental psychology. Human perception and performance*, vol. 5, no. 1, pp. 78–87, 1979. DOI: 10.1037//0096-1523.5.1.78.

[249] S. R. Ellis, S. Smith, and M. W. McGreevy, "Distortions of perceived visual out of pictures", *Perception & Psychophysics*, vol. 42, no. 6, pp. 535–544, Nov. 1, 1987. DOI: 10.3758/BF03207985.

[250] A. Wohlschläger and A. Wohlschläger, "Mental and manual rotation", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 24, no. 2, pp. 397–412, 1998. DOI: 10.1037/0096-1523.24.2.397.

[251] P. Khooshabeh, M. Hegarty, and T. Shipley, "Individual differences in mental rotation", *Experimental Psychology*, vol. 60, pp. 1–8, Nov. 2012. DOI: 10.1027/1618-3169/a000184.

[252] A. Nazareth, R. Killick, A. S. Dick, and S. M. Pruden, "Strategy selection versus flexibility: Using eye-trackers to investigate strategy use during mental rotation", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 45, no. 2, pp. 232–245, 2019. DOI: 10.1037/xlm0000574.

[253] Z. W. Pylyshyn, "What the mind's eye tells the mind's brain: A critique of mental imagery", *Psychological Bulletin*, vol. 80, no. 1, pp. 1–24, 1973. DOI: 10.1037/h0034650.

[254] A. Larsen, "Deconstructing mental rotation", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 40, no. 3, pp. 1072–1091, 2014. DOI: 10.1037/a0035648.

[255] C. Scheer, F. Mattioni Maturana, and P. Jansen, "Sex differences in a chronometric mental rotation test with cube figures: A behavioral, electroencephalography, and eye-tracking pilot study", *NeuroReport*, vol. 29, no. 10, 2018. DOI: 10.1097/WNR.0000000000001046.

[256] S. Fitzhugh, T. Shipley, N. Newcombe, K. McKenna, and D. Dumay, "Mental rotation of real word shepard-metzler figures: An eye tracking study", *Journal of Vision*, vol. 8, pp. 648–648, 2010. DOI: 10.1167/8.6.648.

[257] C. de'Sperati, "Saccades to mentally rotated targets", *Experimental brain research*, vol. 126, no. 4, pp. 563–577, Jun. 1, 1999. DOI: 10.1007/s002210050765.

[258] K. Rayner, "Eye movements in reading and information processing: 20 years of research.", *Psychological Bulletin*, vol. 124, no. 3, pp. 372–422, 1998. DOI: 10.1037/0033-2909.124.3.372.

[259] P. Khooshabeh and M. Hegarty, "Representations of shape during mental rotation.", *AAAI Spring Symposium: Cognitive Shape Processing*, Jan. 1, 2010.

[260] S. T. Iqbal, X. S. Zheng, and B. P. Bailey, "Task-evoked pupillary response to mental workload in human-computer interaction", in *Extended abstracts of the 2004 conference on Human factors and computing systems*, ser. CHI '04, Vienna, Austria: ACM, 2004, p. 1477. DOI: 10.1145/985921.986094.

[261] G. Aston-Jones and J. D. Cohen, "An integrative theory of locus coeruleus- norepinephrine function: Adaptive gain and optimal performance", *Annual review of neuroscience*, vol. 28, pp. 403–450, 2005. DOI: 10.1146/annurev.neuro.28.061604.135709.

[262] W. X. Chmielewski, M. Mückschel, T. Ziemssen, and C. Beste, "The norepinephrine system affects specific neurophysiological subprocesses in the modulation of inhibitory control by working memory demands", *Human Brain Mapping*, vol. 38, no. 1, pp. 68–81, 2017. DOI: 10.1002/hbm.23344.

[263] M. S. Gilzenrat, S. Nieuwenhuis, M. Jepma, and J. D. Cohen, "Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function", *Cognitive, Affective, & Behavioral Neuroscience*, vol. 10, no. 2, pp. 252–269, Jun. 1, 2010. DOI: 10.3758/CABN.10.2.252.

[264] H. M. Rao, R. Khanna, D. J. Zielinski, Y. Lu, J. M. Clements, N. D. Potter, M. A. Sommer, R. Kopper, and L. G. Appelbaum, "Sensorimotor learning during a marksmanship task in immersive virtual reality", *Frontiers in Psychology*, vol. 9, 2018. DOI: 10.3389/fpsyg.2018.00058.

[265] J. H. Friedman, "Stochastic gradient boosting", *Computational Statistics & Data Analysis*, Nonlinear Methods and Data Mining, vol. 38, no. 4, pp. 367–378, Feb. 28, 2002. DOI: 10.1016/S0167-9473(01)00065-2.

[266] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.

[267] P. Jansen, A. Render, C. Scheer, and M. Siebertz, "Mental rotation with abstract and embodied objects as stimuli: Evidence from event-related potential (ERP)", *Experimental Brain Research*, vol. 238, no. 3, pp. 525–535, 2020. DOI: 10.1007/s00221-020-05734-w.

[268] H. Kawamichi, Y. Kikuchi, and S. Ueno, "Spatio-temporal brain activity related to rotation method during a mental rotation task of three-dimensional objects: An MEG study", *NeuroImage*, vol. 37, no. 3, pp. 956–965, Oct. 2007. DOI: 10.1016/j.neuroimage.2007.06.001.

[269] A. T. Bahill, M. R. Clark, and L. Stark, "The main sequence, a tool for studying human eye movements", *Mathematical Biosciences*, vol. 24, no. 3, pp. 191–204, 1975. DOI: 10.1016/0025-5564(75)90075-9.

[270] M. Nakayama, K. Takahashi, and Y. Shimizu, "The act of task difficulty and eye-movement frequency for the 'oculo-motor indices'", in *Proceedings of the 2002 symposium on Eye tracking research & applications*, ser. ETRA '02, ACM, New York, 2002, pp. 37–42. DOI: 10.1145/507072.507080.

[271] J. Goumans, M. M. J. Houben, J. Dits, and J. v. d. Steen, "Peaks and troughs of three-dimensional vestibulo-ocular reflex in humans", *Journal of the Association for Re-*

search in Otolaryngology, vol. 11, no. 3, pp. 383–393, 2010. DOI: 10.1007/s10162-010-0210-y.

[272] R. Allison, M. Eizenman, and B. Cheung, "Combined head and eye tracking system for dynamic testing of the vestibular system", *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 11, pp. 1073–1082, 1996. DOI: 10.1109/10.541249.

[273] E. Games. "Unreal engine, version 4.23.1". (2019), [Online]. Available: https://www.unrealengine.com (visited on 05/13/2024).

[274] M. Peters, B. Laeng, K. Latham, M. Jackson, R. Zaiyouna, and C. Richardson, "A redrawn vandenberg and kuse mental rotations test - different versions and factors that affect performance", *Brain and Cognition*, vol. 28, no. 1, pp. 39–58, Jun. 1995. DOI: 10.1006/brcg.1995.1032.

[275] L. A. Burton and D. Henninger, "Sex differences in relationships between verbal fluency and personality", *Current Psychology*, vol. 32, no. 2, pp. 168–174, Jun. 1, 2013. DOI: 10.1007/s12144-013-9167-4.

[276] M. Hegarty, "Ability and sex differences in spatial thinking: What does the mental rotation test really measure?", *Psychonomic Bulletin & Review*, vol. 25, no. 3, pp. 1212–1219, Jun. 2018. DOI: 10.3758/s13423-017-1347-z.

[277] A. Caissie, F. Vigneau, and D. Bors, "What does the mental rotation test measure? an analysis of item difficulty and item characteristics", *The Open Psychology Journal*, vol. 2, pp. 94–102, 2009. DOI: 10.2174/1874350100902010094.

[278] N. S. Holliman, A. Coltekin, S. J. Fernstad, L. McLaughlin, M. D. Simpson, and A. J. Woods, "Visual entropy and the visualization of uncertainty", *arXiv:1907.12879*, Apr. 2022.

[279] H. R. Schiffman, *Sensation and Perception: An Integrated Approach*. New York: John Wiley & Sons, Jan. 15, 2001, 608 pp.

[280] S. Mathôt and A. Vilotijević, "Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis", *Behavior Research Methods*, vol. 55, pp. 3055–3077, Aug. 26, 2022. DOI: 10.3758/s13428-022-01957-7.

[281] M. E. Kret and E. E. Sjak-Shie, "Preprocessing pupil size data: Guidelines and code", *Behavior Research Methods*, vol. 51, no. 3, pp. 1336–1342, 2019. DOI: 10.3758/s13428-018-1075-y.

[282] S. Mathôt, J. Fabius, E. Van Heusden, and S. Van der Stigchel, "Safe and sensible preprocessing and baseline correction of pupil-size data", *Behavior Research Methods*, vol. 50, no. 1, pp. 94–106, 2018. DOI: 10.3758/s13428-017-1007-2.

[283] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols", in *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, New York, NY, USA: ACM, 2000, pp. 71–78. DOI: 10.1145/355017.355028.

[284] I. Agtzidis, M. Startsev, and M. Dorr, "360-degree video gaze behaviour: A ground-truth data set and a classification algorithm for eye movements", in *Proceedings of the 27th ACM International Conference on Multimedia*, New York, NY, USA: ACM, 2019, pp. 1007–1015. DOI: 10.1145/3343031.3350947.

[285] R. Andersson, M. Nyström, and K. Holmqvist, "Sampling frequency and eye-tracking measures: How speed affects durations, latencies, and more", *Journal of Eye Movement Research*, vol. 3, no. 3, Sep. 13, 2010. DOI: 10.16910/jemr.3.3.6.

[286] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in python", *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.

[287] T. Appel, C. Scharinger, P. Gerjets, and E. Kasneci, "Cross-subject workload classification using pupil-related measures", in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, ser. ETRA '18, New York, NY, USA: ACM, Jun. 14, 2018, pp. 1–8. DOI: 10.1145/3204493.3204531.

[288] T. Appel, P. Gerjets, S. Hoffman, K. Moeller, M. Ninaus, C. Scharinger, N. Sevcenko, F. Wortha, and E. Kasneci, "Cross-task and cross-participant classification of cognitive load in an emergency simulation game", *IEEE Transactions on Affective Computing*, vol. 14, no. 2, pp. 1558–1571, 2021. DOI: 10.1109/TAFFC.2021.3098237.

[289] M. V. Alves, S. Tassini, F. Aedo-Jury, and O. F. A. Bueno, *Cognitive processing dissociation by mental effort manipulation in long demanding tasks*, Apr. 27, 2020. DOI: 10.1101/2020.04.25.060814.

[290] S. Mathôt, *Python DataMatrix*, Aug. 26, 2022.

[291] S. Mathôt, "A simple way to reconstruct pupil size during eye blinks", *FigShare*, Apr. 24, 2013. DOI: 10.6084/m9.figshare.688001.

[292] J. Hadnett-Hunter, G. Nicolaou, E. O'Neill, and M. Proulx, "The effect of task on visual attention in interactive virtual environments", *ACM Trans. Appl. Percept.*, vol. 16, no. 3, 2019. DOI: 10.1145/3352763.

[293] S. Kavanagh, A. Luxton-Reilly, B. Wuensche, and B. Plimmer, "A systematic review of virtual reality in education", *Themes in Science and Technology Education*, vol. 10, no. 2, pp. 85–119, 2017.

[294] L. Freina and M. Ott, "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*, Bucharest, Romania: Carol I NDU Publishing House, 2015, pp. 133–141. DOI: 10.12753/2066-026X-15-020.

[295] C. Moro, Z. Štromberga, A. Raikos, and A. Stirling, "The effectiveness of virtual and augmented reality in health sciences and medical anatomy", *Anatomical Sciences Education*, vol. 10, no. 6, pp. 549–559, 2017. DOI: 10.1002/ase.1696.

[296] W. Alhalabi, "Virtual reality systems enhance students' achievements in engineering education", *Behaviour & Information Technology*, vol. 35, no. 11, pp. 919–925, Nov. 1, 2016. DOI: https://doi.org/10.1080/0144929X.2016.1212931.

[297] A. Casu, L. D. Spano, F. Sorrentino, and R. Scateni, "RiftArt: Bringing masterpieces in the classroom through immersive virtual reality", in *Smart tools and apps for graphics - eurographics italian chapter conference*, Geneva, Switzerland: The Eurographics Association, 2015, pp. 77–84. DOI: 10.2312/stag.20151294.

[298] K.-H. Cheng and C.-C. Tsai, "A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors", *Computers & Education*, vol. 140, p. 103 600, Oct. 1, 2019. DOI: 10.1016/j.compedu.2019.103600.

[299] R. Lamb and E. A. Etopio, "Virtual reality: A tool for preservice science teachers to put theory into practice", *Journal of Science Education and Technology*, vol. 29, no. 4, pp. 573–585, 2020. DOI: 10.1007/s10956-020-09837-5.

[300] A. L. Simeone, M. Speicher, A. Molnar, A. Wilde, and F. Daiber, "LIVE: The human role in learning in immersive virtual environments", in *Symposium on spatial user interaction*, New York, NY, USA: ACM, 2019. DOI: 10.1145/3357251.3357590.

[301] M.-Y. Liao, C.-Y. Sung, H.-C. Wang, and W.-C. Lin, "Virtual classmates: Embodying historical learners' messages as learning companions in a VR classroom through comment mapping", in *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*, New York, NY, USA: IEEE, 2019, pp. 163–171. DOI: 10.1109/VR.2019.8797708.

[302] S. Sharma, R. Agada, and J. Ruffin, "Virtual reality classroom as an constructivist approach", in *2013 Proceedings of IEEE Southeastcon*, Jacksonville, FL, USA: IEEE, 2013, pp. 1–5. DOI: 10.1109/SECON.2013.6567441.

[303] W. Fuhl, M. Tonsen, A. Bulling, and E. Kasneci, "Pupil detection for head-mounted eye tracking in the wild: An evaluation of the state of the art", *Machine Vision and Applications*, vol. 27, no. 8, 2016. DOI: 10.1007/s00138-016-0776-4.

[304] X. Zhang, S. Park, T. Beeler, D. Bradley, S. Tang, and O. Hilliges, "ETH-XGaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation", in *Computer vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., Cham, Switzerland: Springer International Publishing, 2020, pp. 365–381. DOI: 10.1007/978-3-030-58558-7_22.

[305] A. Schmitz, A. MacQuarrie, S. Julier, N. Binetti, and A. Steed, "Directing versus attracting attention: Exploring the effectiveness of central and peripheral cues in panoramic videos", in *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*, New York, NY, USA: IEEE, 2020, pp. 63–72. DOI: 10.1109/VR46266.2020.00024.

[306] T. M. Lee, J.-C. Yoon, and I.-K. Lee, "Motion sickness prediction in stereoscopic videos using 3d convolutional neural networks", *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 1919–1927, 2019. DOI: 10.1109/TVCG.2019.2899186.

[307] Y. Lang, L. Wei, F. Xu, Y. Zhao, and L.-F. Yu, "Synthesizing personalized training programs for improving driving habits via virtual reality", in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Tuebingen/Reutlingen, Germany: IEEE, 2018, pp. 297–304. DOI: 10.1109/VR.2018.8448290.

[308] E. Arabadzhiyska, O. T. Tursun, K. Myszkowski, H.-P. Seidel, and P. Didyk, "Saccade landing position prediction for gaze-contingent rendering", *ACM Trans. Graph.*, vol. 36, no. 4, 2017. DOI: 10.1145/3072959.3073642.

[309] X. Meng, R. Du, and A. Varshney, "Eye-dominance-guided foveated rendering", *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 5, pp. 1972–1980, 2020. DOI: 10.1109/TVCG.2020.2973442.

[310] J. Orlosky, Y. Itoh, M. Ranchet, K. Kiyokawa, J. Morgan, and H. Devos, "Emulation of physician tasks in eye-tracked virtual reality for remote diagnosis of neurodegenerative disease", *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, pp. 1302–1311, 2017. DOI: 10.1109/TVCG.2017.2657018.

[311] E. Langbehn, F. Steinicke, M. Lappe, G. F. Welch, and G. Bruder, "In the blink of an eye: Leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality", *ACM Trans. Graph.*, vol. 37, no. 4, 2018. DOI: 10.1145/3197517.3201335.

[312] Y. Zhang, W. Hu, W. Xu, C. T. Chou, and J. Hu, "Continuous authentication using eye movement response of implicit visual stimuli", *ACM Interact. Mob. Wearable Ubiquitous Technology*, vol. 1, no. 4, 177:1–177:22, 2018. DOI: 10.1145/3161410.

[313] E. Bozkir, D. Geisler, and E. Kasneci, "Assessment of driver attention during a safety critical situation in VR to generate VR-based training", in *ACM symposium on applied perception 2019*, New York, NY, USA: ACM, 2019. DOI: 10.1145/3343036.3343138.

[314] M. Khamis, C. Oechsner, F. Alt, and A. Bulling, "VRPursuits: Interaction in virtual reality using smooth pursuit eye movements", in *Proceedings of the 2018 international conference on advanced visual interfaces*, New York, NY, USA: ACM, 2018. DOI: 10.1145/3206505.3206522.

[315] L. Sidenmark and A. Lundström, "Gaze behaviour on interacted objects during hand interaction in virtual reality for eye tracking calibration", in *Proceedings of the 11th ACM symposium on eye tracking research & applications*, New York, NY, USA: ACM, 2019. DOI: 10.1145/3314111.3319815.

[316] D. Weintrop, E. Beheshti, M. Horn, O. Kai, K. Jona, L. Trouille, and U. Wilensky, "Defining computational thinking for mathematics and science classrooms", *Journal of Science Education and Technology*, vol. 25, no. 1, pp. 127–147, 2016. DOI: 10.1007/s10956-015-9581-5.

[317] M. Lombard, T. Bolmarcich, and L. Weinstein, "Measuring presence: The temple presence inventory", in *Proceedings of the 12th annual international workshop on*

presence, Los Angeles, CA, USA: The International Society for Presence Research, 2009, pp. 1–15.

[318] S. D. Roth, "Ray casting for modeling solids", *Computer Graphics and Image Processing*, vol. 18, no. 2, pp. 109–144, 1982. DOI: 10.1016/0146-664X(82)90169-1.

[319] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only anova procedures", in *Proceedings of the SIGCHI conference on human factors in computing systems*, New York, NY, USA: ACM, 2011, pp. 143–146. DOI: 10.1145/1978942.1978963.

[320] H. W. Marsh and J. W. Parker, "Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well?", *Journal of Personality and Social Psychology*, vol. 47, no. 1, pp. 213–231, 1984. DOI: 10.1037/0022-3514.47.1.213.

[321] E. B. Cloude, D. A. Dever, M. D. Wiedbusch, and R. Azevedo, "Quantifying scientific thinking using multichannel data with crystal island: Implications for individualized game-learning analytics", *Frontiers in Education*, vol. 5, p. 217, 2020. DOI: 10.3389/feduc.2020.572546.

[322] M. D. Wiedbusch and R. Azevedo, "Modeling metacomprehension monitoring accuracy with eye gaze on informational content in a multimedia learning environment", in *ACM Symposium on Eye Tracking Research and Applications*, ser. ETRA '20 Full Papers, vol. 20, New York, NY, USA: ACM, 2020. DOI: 10.1145/3379155.3391329.

[323] Ö. Sümer, P. Gerjets, U. Trautwein, and E. Kasneci, "Automated anonymisation of visual and audio data in classroom studies", in *The workshops of the thirty-forth AAAI conference on artificial intelligence*, Palo Alto, CA, USA: AAAI Press, 2020.

[324] A. K. Chaudhary and J. B. Pelz, "Privacy-preserving eye videos using rubber sheet model", in *ACM Symposium on Eye Tracking Research and Applications*, ser. ETRA '20 Short Papers, vol. 22, New York, NY, USA: ACM, 2020. DOI: 10.1145/3379156.3391375.

[325] R. Böheim, T. Urdan, M. Knogler, and T. Seidel, "Student hand-raising as an indicator of behavioral engagement and its role in classroom learning", *Contemporary Educational Psychology*, vol. 62, p. 101 894, Jul. 1, 2020. DOI: 10.1016/j.cedpsych.2020.101894.

[326] P. Nahlik and P. L. Daubenmire, "Adapting gaze-transition entropy analysis to compare participants' problem solving approaches for chemistry word problems", *Chemistry Education Research and Practice*, vol. 23, no. 3, pp. 714–724, 2022. DOI: 10.1039/D2RP00066K.

[327] S. Spencer, T. Drescher, J. Sears, A. Scruggs, and J. Schreffler, "Comparing the efficacy of virtual simulation to traditional classroom role-play", *Journal of Educational Computing Research*, vol. 57, no. 7, pp. 1772–1785, Jun. 25, 2019. DOI: 10.1177/0735633119855613.

[328] J. O. Bailey and J. N. Bailenson, "Chapter 9 - immersive virtual reality and the developing child", in *Cognitive Development in Digital Contexts*, F. C. Blumberg and P. J. Brooks, Eds., San Diego: Academic Press, Jan. 1, 2017, pp. 181–200. DOI: 10.1016/B978-0-12-809481-5.00009-2.

[329] F. Shic, J. Bradshaw, A. Klin, B. Scassellati, and K. Chawarska, "Limited activity monitoring in toddlers with autism spectrum disorder", *Brain Research*, vol. 1380, pp. 246–254, Mar. 22, 2011. DOI: 10.1016/j.brainres.2010.11.074.

[330] M. Mikhailenko, N. Maksimenko, and M. Kurushkin, "Eye-tracking in immersive virtual reality for education: A review of the current progress and applications", *Frontiers in Education*, vol. 7, 2022. DOI: https://doi.org/10.3389/feduc.2022.697032.

[331] B. A. Shiferaw, L. A. Downey, J. Westlake, B. Stevens, S. M. W. Rajaratnam, D. J. Berlowitz, P. Swann, and M. E. Howard, "Stationary gaze entropy predicts lane departure events in sleep-deprived drivers", *Scientific Reports*, vol. 8, no. 1, p. 2220, Feb. 2, 2018. DOI: 10.1038/s41598-018-20588-7.

[332] K. Krejtz, A. T. Duchowski, K. Wisiecka, and I. Krejtz, "Entropy of eye movements while reading code or text", in *Proceedings of the Tenth International Workshop on Eye Movements in Programming*, ser. EMIP '22, New York, NY, USA: ACM, Nov. 28, 2022, pp. 8–14. DOI: 10.1145/3524488.3527365.

[333] I. A. Ebeid and J. Gwizdka, "Real-time gaze transition entropy", in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, ser. ETRA '18, New York, NY, USA: ACM, Jun. 14, 2018, pp. 1–3. DOI: 10.1145/3204493.3208340.

[334] Y. Huang, E. Richter, T. Kleickmann, A. Wiepke, and D. Richter, "Classroom complexity affects student teachers' behavior in a VR classroom", *Computers & Education*, vol. 163, 2020. DOI: https://doi.org/10.1016/j.compedu.2020.104100.

[335] Tobii. "Eye tracking technology for VR - VIVE pro eye with tobii". (), [Online]. Available: https://www.tobii.com/products/integration/xr-headsets/device-integrations/htc-vive-pro-eye (visited on 02/26/2024).

[336] J. Golbeck, "Chapter 3 - network structure and measures", in *Analyzing the Social Web*, J. Golbeck, Ed., Boston: Morgan Kaufmann, Jan. 1, 2013, pp. 25–44. DOI: 10.1016/B978-0-12-405531-5.00003-1.

[337] O. Serrat, "Social network analysis", in *Knowledge Solutions: Tools, Methods, and Approaches to Drive Organizational Performance*, O. Serrat, Ed., Singapore: Springer, 2017, pp. 39–43. DOI: 10.1007/978-981-10-0983-9_9.

[338] T. A. B. Snijders, "Social network analysis", in *International Encyclopedia of Statistical Science*, M. Lovric, Ed., Berlin, Heidelberg: Springer, 2011, pp. 1356–1358.

[339] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Structural Analysis in the Social Sciences). Cambridge: Cambridge University Press, 1994. DOI: 10.1017/CBO9780511815478.

[340] K. Erciyes, *Discrete Mathematics and Graph Theory. A Concise Study Companion and Guide*. Springer, Cham, 2021.

[341] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths", *Social Networks*, vol. 32, no. 3, pp. 245–251, Jul. 1, 2010. DOI: 10.1016/j.socnet.2010.03.006.

[342] S. Pastel, C.-H. Chen, L. Martin, M. Naujoks, K. Petri, and K. Witte, "Comparison of gaze accuracy and precision in real-world and virtual reality", *Virtual Reality*, vol. 25, no. 1, pp. 175–189, 2021. DOI: 10.1007/s10055-020-00449-3.

[343] S. Weber, R. S. Schubert, S. Vogt, B. M. Velichkovsky, and S. Pannasch, "Gaze3dfix: Detecting 3d fixations with an ellipsoidal bounding volume", *Behavior Research Methods*, vol. 50, no. 5, pp. 2004–2015, Oct. 1, 2018. DOI: 10.3758/s13428-017-0969-4.

[344] K. Holmqvist, S. L. Örbom, I. T. C. Hooge, D. C. Niehorster, R. G. Alexander, R. Andersson, J. S. Benjamins, P. Blignaut, A.-M. Brouwer, L. L. Chuang, K. A. Dalrymple, D. Drieghe, M. J. Dunn, U. Ettinger, S. Fiedler, T. Foulsham, J. N. van der Geest, D. W. Hansen, S. B. Hutton, E. Kasneci, A. Kingstone, P. C. Knox, E. M. Kok, H. Lee, J. Y. Lee, J. M. Leppänen, S. Macknik, P. Majaranta, S. Martinez-Conde, A. Nuthmann, M. Nyström, J. L. Orquin, J. Otero-Millan, S. Y. Park, S. Popelka, F. Proudlock, F. Renkewitz,

A. Roorda, M. Schulte-Mecklenbeck, B. Sharif, F. Shic, M. Shovman, M. G. Thomas, W. Venrooij, R. Zemblys, and R. S. Hessels, "Eye tracking: Empirical foundations for a minimal reporting guideline", *Behavior Research Methods*, vol. 55, pp. 364–416, Apr. 6, 2022. DOI: 10.3758/s13428-021-01762-8.

[345] A. T. Duchowski, "Head-mounted system software development", in *Eye Tracking Methodology: Theory and Practice*, A. T. Duchowski, Ed., Cham: Springer International Publishing, 2017, pp. 67–84. DOI: 10.1007/978-3-319-57883-5_7.

[346] HTC. "Eye and facial tracking SDK (legacy) - developer resources". (), [Online]. Available: https://developer-express.vive.com/resources/vive-sense/eye-and-facial-tracking-sdk/ (visited on 02/26/2024).

[347] K. Emperore and D. Sherry, *Unreal Engine Physics Essentials*. Birmingham: Packt Publishing, 2015.

[348] J. Hu and Y. Zhang, "Discovering the interdisciplinary nature of big data research through social network analysis and visualization", *Scientometrics*, vol. 112, no. 1, pp. 91–109, Jul. 1, 2017. DOI: 10.1007/s11192-017-2383-1.

[349] H. Thimbleby and J. Gow, "Applying graph theory to interaction design", in *Engineering Interactive Systems: EIS 2007 Joint Working Conferences, EHCI 2007, DSV-IS 2007, HCSE 2007, Salamanca, Spain, March 22-24, 2007. Selected Papers*, Berlin, Heidelberg: Springer-Verlag, 2008, pp. 501–519.

[350] S. Werner, B. Krieg-Brückner, and T. Herrmann, "Modelling navigational knowledge by route graphs", in *Spatial Cognition II*, C. Freksa, C. Habel, W. Brauer, and K. F. Wender, Eds., vol. 1849, Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 295–316. DOI: 10.1007/3-540-45460-8_22.

[351] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using NetworkX", in *Proceedings of the 7th python in science conference*, G. Varoquaux, T. Vaught, and J. Millman, Eds., Pasadena, CA USA, 2008, pp. 11–15.

[352] L. C. Freeman, "The development of social network analysis – with an emphasis on recent events", in *The SAGE Handbook of Social Network Analysis*. London: SAGE Publications Ltd, Apr. 25, 2023. DOI: 10.4135/9781446294413.

[353] P. Wills and F. G. Meyer, "Metrics for graph comparison: A practitioner's guide", *PLOS ONE*, vol. 15, no. 2, pp. 1–54, Feb. 2020. DOI: 10.1371/journal.pone.0228728.

[354] K. Das, S. Samanta, and M. Pal, "Study on centrality measures in social networks: A survey", *Social Network Analysis and Mining*, vol. 8, no. 1, p. 13, Dec. 2018. DOI: 10.1007/s13278-018-0493-2.

[355] A. Lancichinetti, S. Fortunato, and J. Kertész, "Detecting the overlapping and hierarchical community structure in complex networks", *New Journal of Physics*, vol. 11, no. 3, p. 033 015, Mar. 2009. DOI: 10.1088/1367-2630/11/3/033015.

[356] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental algorithms for scientific computing in python", *Nature Methods*, vol. 17, pp. 261–272, 2020. DOI: 10.1038/s41592-019-0686-2.

[357] M. H. u. Rehman, C. S. Liew, A. Abbas, P. P. Jayaraman, T. Y. Wah, and S. U. Khan, "Big data reduction methods: A survey", *Data Science and Engineering*, vol. 1, no. 4, pp. 265–284, Dec. 1, 2016. DOI: 10.1007/s41019-016-0022-0.

[358] W. Weimer and G. C. Necula, "Finding and preventing run-time error handling mistakes", in *Proceedings of the 19th annual ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications*, ser. OOPSLA '04, New York, NY, USA: ACM, 2004, pp. 419–431. DOI: 10.1145/1028976.1029011.

[359] B. Schiller, C. Deusser, J. Castrillon, and T. Strufe, "Compile- and run-time approaches for the selection of efficient data structures for dynamic graph analysis", *Applied Network Science*, vol. 1, no. 1, p. 9, Sep. 5, 2016. DOI: 10.1007/s41109-016-0011-2.

[360] C. Ge, Y. Li, E. Eilebrecht, B. Chandramouli, and D. Kossmann, "Speculative distributed CSV data parsing for big data analytics", in *Proceedings of the 2019 International Conference on Management of Data*, ser. SIGMOD '19, New York, NY, USA: ACM, Jun. 25, 2019, pp. 883–899. DOI: 10.1145/3299869.3319898.

[361] J. Patty and E. Penn, "Analyzing big data: Social choice and measurement", *Political Science and Politics*, vol. 48, no. 1, pp. 95–101, Jan. 1, 2015. DOI: 10.1017/S1049096514001814.

[362] V. S. Rawat, "Chapter 12 release memory", in *Best Coding Practices for R*, 2022.

[363] B. Johnson and D. A. S. Chandran, "COMPARISON BETWEEN PYTHON, JAVA AND r PROGRMMING LANGUAGE IN MACHINE LEARNING", *International Research Journal of Modernization in Engineering Technology and Science*, vol. 3, no. 6, Jun. 26, 2021.

[364] M. Fey and J. E. Lenssen, *Fast graph representation learning with PyTorch geometric*, Apr. 25, 2019. DOI: 10.48550/arXiv.1903.02428.

[365] G. Csárdi, T. Nepusz, K. Müller, S. Horvát, V. Traag, F. Zanini, and D. Noom, *Igraph: Network analysis and visualization in r*, version R package version 1.6.0, Dec. 13, 2023. DOI: 10.5281/zenodo.10369053.

[366] G. H. Nguyen, J. B. Lee, R. A. Rossi, N. K. Ahmed, E. Koh, and S. Kim, "Continuous-time dynamic network embeddings", in *Companion Proceedings of the The Web Conference 2018*, ser. WWW '18, Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, Apr. 23, 2018, pp. 969–976. DOI: 10.1145/3184558.3191526.

[367] Y. Wang, Y. Yuan, Y. Ma, and G. Wang, "Time-dependent graphs: Definitions, applications, and algorithms", *Data Science and Engineering*, vol. 4, no. 4, pp. 352–366, Dec. 1, 2019. DOI: 10.1007/s41019-019-00105-0.

[368] R. Webster, "Declarative knowledge acquisition in immersive virtual learning environments", *Interactive Learning Environments*, vol. 24, no. 6, Aug. 17, 2015.

[369] J. D. Bric, D. C. Lumbard, M. J. Frelich, and J. C. Gould, "Current state of virtual reality simulation in robotic surgery training: A review", *Surgical Endoscopy*, vol. 30, no. 6, pp. 2169–2178, Jun. 1, 2016. DOI: 10.1007/s00464-015-4517-y.

[370] G. Fauville, A. C. M. Queiroz, L. Hambrick, B. A. Brown, and J. N. Bailenson, "Participatory research on using virtual reality to teach ocean acidification: A study in the marine education community", *Environmental Education Research*, vol. 27, no. 2, pp. 254–278, Sep. 2020. DOI: 10.1080/13504622.2020.1803797.

[371] A. C. M. Queiroz, A. M. Nascimento, R. Tori, and M. I. da Silva Leme, "Using HMD-based immersive virtual environments in primary/k-12 education", in *Immersive Learning Research Network*, D. Beck, C. Allison, L. Morgado, J. Pirker, A. Peña-Rios, T. Ogle, J. Richter, and C. Gütl, Eds., Cham: Springer International Publishing, 2018, pp. 160–173. DOI: 10.1007/978-3-319-93596-6_11.

[372] M. C. Howard, "Virtual reality interventions for personal development: A meta-analysis of hardware and software", *Human–Computer Interaction*, vol. 34, no. 3, pp. 205–239, May 4, 2018. DOI: 10.1080/07370024.2018.1469408.

[373] P. A. Alexander, "Past as prologue: Educational psychology's legacy and progeny", *Journal of Educational Psychology*, vol. 110, no. 2, pp. 147–162, 2018. DOI: 10.1037/edu0000200.

[374] J. E. Brophy and T. L. Good, *Teacher-student relationships: Causes and consequences* (Teacher-student relationships: Causes and consequences). Oxford, England: Holt, Rinehart & Winston, 1974, xvi, 400.

[375] R. Harker and P. Tymms, "The effects of student composition on school outcomes", *School Effectiveness and School Improvement*, vol. 15, no. 2, pp. 177–199, Jun. 1, 2004. DOI: 10.1076/sesi.15.2.177.30432.

[376] H. W. Marsh, O. Lüdtke, B. Nagengast, U. Trautwein, A. J. S. Morin, A. S. Abduljabbar, and O. Köller, "Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects", *Educational Psychologist*, vol. 47, no. 2, pp. 106–124, Apr. 2012. DOI: 10.1080/00461520.2012.670488.

[377] M. A. Gottfried, "Peer effects in urban schools: Assessing the impact of classroom composition on student achievement", *Educational Policy*, vol. 28, no. 5, pp. 607–647, Dec. 7, 2012. DOI: 10.1177/0895904812467082.

[378] J. A. C. Hattie, "Classroom composition and peer effects", *International Journal of Educational Research*, vol. 37, no. 5, pp. 449–481, Jan. 1, 2002. DOI: 10.1016/S0883-0355(03)00015-6.

[379] J. Hochweber, I. Hosenfeld, and E. Klieme, "Classroom composition, classroom management, and the relationship between student attributes and grades", *Journal of Educational Psychology*, vol. 106, no. 1, pp. 289–300, 2014. DOI: 10.1037/a0033829.

[380] V. Lavy, M. D. Paserman, and A. Schlosser, "Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom", *The Economic Journal*, vol. 122, no. 559, pp. 208–237, Mar. 1, 2012. DOI: 10.1111/j.1468-0297.2011.02463.x.

[381] P. L. Hardré and D. W. Sullivan, "Student differences and environment perceptions: How they contribute to student motivation in rural high schools", *Learning and Individual Differences*, Including Special Issue on Creativity, vol. 18, no. 4, pp. 471–485, Oct. 1, 2008. DOI: 10.1016/j.lindif.2007.11.010.

[382] R. Pekrun, K. Murayama, H. W. Marsh, T. Goetz, and A. C. Frenzel, "Happy fish in little ponds: Testing a reference group model of achievement and emotion", *Journal of Personality and Social Psychology*, vol. 117, no. 1, pp. 166–185, 2019. DOI: 10.1037/pspp0000230.

[383] P. Dijkstra, H. Kuyper, G. van der Werf, A. P. Buunk, and Y. G. van der Zee, "Social comparison in the classroom: A review", *Review of Educational Research*, vol. 78, no. 4, pp. 828–879, Dec. 1, 2008. DOI: 10.3102/0034654308321210.

[384] U. Trautwein, H. Dumont, and A.-L. Dicke, "Schooling: Impact on cognitive and motivational development", in *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, J. D. Wright, Ed., Oxford: Elsevier, Jan. 1, 2015, pp. 119–124. DOI: 10.1016/B978-0-08-097086-8.26056-X.

[385] J. Fox, D. Arena, and J. N. Bailenson, "Virtual reality", *Journal of Media Psychology*, vol. 21, no. 3, pp. 95–113, Jan. 2009. DOI: 10.1027/1864-1105.21.3.95.

[386] I. Hudson and J. Hurter, "Avatar types matter: Review of avatar literature for performance purposes", in *Virtual, Augmented and Mixed Reality*, S. Lackey and R. Shumaker, Eds., Cham: Springer International Publishing, 2016, pp. 14–21. DOI: 10.1007/978-3-319-39907-2_2.

[387] T. Mussweiler, "Comparison processes in social judgment: Mechanisms and consequences", *Psychological Review*, vol. 110, no. 3, pp. 472–489, 2003. DOI: 10.1037/0033-295X.110.3.472.

[388] J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher, and M. S. Wetherell, *Rediscovering the social group: A self-categorization theory* (Rediscovering the social group: A self-categorization theory). Cambridge, MA, US: Basil Blackwell, 1987, x, 239.

[389] D. J. MacAulay, "Classroom environment: A literature review", *Educational Psychology*, vol. 10, no. 3, pp. 239–253, 1990. DOI: 10.1080/0144341900100305.

[390] R. Wannarka and K. Ruhl, "Seating arrangements that promote positive academic and behavioural outcomes: A review of empirical research", *Support for Learning*, vol. 23, no. 2, pp. 89–93, 2008. DOI: 10.1111/j.1467-9604.2008.00375.x.

[391] A. C. Fernandes, J. Huang, and V. Rinaldo, "Does Where A Student Sits Really Matter? - The Impact of Seating Locations on Student Classroom Learning", *International Journal of Applied Educational Studies*, vol. 10, no. 1, 2011.

[392] K. Lacroix and S. Lacroix, "Does seat location matter? a review of the proximity effect in large and small classrooms", *Community College Enterprise*, vol. 23, no. 2, Jan. 1, 2017.

[393] D. W. Levine, E. C. O'Neal, S. G. Garwood, and P. J. McDonald, "Classroom ecology: The effects of seating position on grades and participation", *Personality and Social Psychology Bulletin*, vol. 6, no. 3, pp. 409–412, Sep. 1, 1980. DOI: 10.1177/014616728063012.

[394] M. D. Meeks, T. L. Knotts, K. D. James, F. Williams, J. A. Vassar, and A. O. Wren, "The impact of seating location and seating type on student performance", *Education Sciences*, vol. 3, no. 4, pp. 375–386, Dec. 2013. DOI: 10.3390/educsci3040375.

[395] D. R. Montello, "Classroom seating location and its effect on course achievement, participation, and attitudes", *Journal of Environmental Psychology*, vol. 8, no. 2, pp. 149–157, 1988. DOI: 10.1016/S0272-4944(88)80005-7.

[396] K. K. Perkins and C. E. Wieman, "The surprising impact of seat location on student performance", *The Physics Teacher*, vol. 43, no. 1, pp. 30–33, Jan. 1, 2005. DOI: 10.1119/1.1845987.

[397] A. I. Schwebel and D. L. Cherlin, "Physical and social distancing in teacher-pupil relationships", *Journal of Educational Psychology*, vol. 63, no. 6, pp. 543–550, 1972. DOI: 10.1037/h0034081.

[398] P. Will, W. F. Bischof, and A. Kingstone, "The impact of classroom seating location and computer use on student academic performance", *PLOS ONE*, vol. 15, no. 8, e0236131, May 8, 2020. DOI: 10.1371/journal.pone.0236131.

[399] L. Cheng, S. Farnham, and L. Stone, "Lessons learned: Building and deploying shared virtual environments", in *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, R. Schroeder, Ed., London: Springer, 2002, pp. 90–111. DOI: 10.1007/978-1-4471-0277-9_6.

[400] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley [from the field]", *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 98–100, Jun. 2012. DOI: 10.1109/MRA.2012.2192811.

[401] C.-C. Ho and K. F. MacDorman, "Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices", *Computers in Human Behavior*, vol. 26, no. 6, pp. 1508–1518, Nov. 1, 2010. DOI: 10.1016/j.chb.2010.05.015.

[402] K. Macdorman, "Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley", in *ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science*, 2006, pp. 25–29.

[403] M. B. Mathur and D. B. Reichling, "Navigating a social world with robot partners: A quantitative cartography of the uncanny valley", *Cognition*, vol. 146, pp. 22–32, Jan. 1, 2016. DOI: 10.1016/j.cognition.2015.09.008.

[404] M. Strait, L. Vujovic, V. Floerke, M. Scheutz, and H. Urry, "Too much humanness for human-robot interaction: Exposure to highly humanlike robots elicits aversive responding in observers", in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15, New York, NY, USA: ACM, Apr. 18, 2015, pp. 3593–3602. DOI: 10.1145/2702123.2702415.

[405] P. Heidicker, E. Langbehn, and F. Steinicke, "Influence of avatar appearance on presence in social VR", in *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, Mar. 2017, pp. 233–234. DOI: 10.1109/3DUI.2017.7893357.

[406] G. Makransky, P. Wismer, and R. E. Mayer, "A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation", *Journal of Computer Assisted Learning*, vol. 35, no. 3, pp. 349–358, 2019. DOI: 10.1111/jcal.12335.

[407] C. Zanbaka, P. Goolkasian, and L. Hodges, "Can a virtual cat persuade you? the role of gender and realism in speaker persuasiveness", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '06, New York, NY, USA: ACM, Apr. 22, 2006, pp. 1153–1162. DOI: 10.1145/1124772.1124945.

[408] M. B. Brewer and J. G. Weber, "Self-evaluation effects of interpersonal versus intergroup social comparison", *Journal of Personality and Social Psychology*, vol. 66, no. 2, pp. 268–275, 1994. DOI: 10.1037/0022-3514.66.2.268.

[409] B. DE Fraine, J. Van Damme, G. Van Landeghem, M.-C. Opdenakker, and P. Onghena, "The effect of schools and classes on language achievement", *British Educational Research Journal*, vol. 29, no. 6, pp. 841–859, 2003. DOI: 10.1080/0141192032000137330.

[410] J. C. Fruehwirth, "Identifying peer achievement spillovers: Implications for desegregation and the achievement gap", *Quantitative Economics*, vol. 4, no. 1, pp. 85–124, 2013. DOI: 10.3982/QE93.

[411] B. W. Pelham and J. O. Wachsmuth, "The waxing and waning of the social self: Assimilation and contrast in social comparison", *Journal of Personality and Social Psychology*, vol. 69, no. 5, pp. 825–838, 1995. DOI: 10.1037/0022-3514.69.5.825.

[412] J. Fang, X. Huang, M. Zhang, F. Huang, Z. Li, and Q. Yuan, "The big-fish-little-pond effect on academic self-concept: A meta-analysis", *Frontiers in Psychology*, vol. 9, Aug. 29, 2018. DOI: 10.3389/fpsyg.2018.01569.

[413] H. W. Marsh, A. J. Martin, A. S. Yeung, and R. G. Craven, "Competence self- perceptions", in *Handbook of competence and motivation: Theory and application, 2nd ed*, New York, NY, US: The Guilford Press, 2017, pp. 85–115.

[414] H. W. Marsh and M. Seaton, "Chapter five - the big-fish–little-pond effect, competence self-perceptions, and relativity: Substantive advances and methodological innovation", in *Advances in Motivation Science*, A. J. Elliot, Ed., vol. 2, Elsevier, Jan. 1, 2015, pp. 127–184. DOI: 10.1016/bs.adms.2015.05.002.

[415] A. Bönsch, J. Wendt, H. Overath, Ö. Gürerk, C. Harbring, C. Grund, T. Kittsteiner, and T. W. Kuhlen, "Peers at work: Economic real-effort experiments in the presence of virtual co-workers", in *2017 IEEE Virtual Reality (VR)*, Los Angeles, CA, USA, Mar. 2017, pp. 301–302. DOI: 10.1109/VR.2017.7892296.

[416] Ö. Gürerk, A. Bönsch, T. Kittsteiner, and A. Staffeldt, "Virtual humans as co-workers: A novel methodology to study peer effects", *Journal of Behavioral and Experimental Economics*, vol. 78, pp. 17–29, Feb. 1, 2019. DOI: 10.1016/j.socec.2018.11.003.

[417] H. Jarodzka, K. Holmqvist, and H. Gruber, "Eye tracking in educational science: Theoretical frameworks and research agendas", *Journal of Eye Movement Research*, vol. 10, no. 1, Feb. 4, 2017. DOI: 10.16910/jemr.10.1.3.

[418] A. R. Strohmaier, K. J. MacKay, A. Obersteiner, and K. M. Reiss, "Eye-tracking methodology in mathematics education research: A systematic literature review", *Educational Studies in Mathematics*, vol. 104, no. 2, pp. 147–200, Jun. 1, 2020. DOI: 10.1007/s10649-020-09948-1.

[419] C. Bundesen, "A theory of visual attention", *Psychological Review*, vol. 97, no. 4, pp. 523–547, 1990. DOI: 10.1037/0033-295X.97.4.523.

[420] J. Lodge and W. Harrison, "The role of attention in learning in the digital age", *The Yale Journal of Biology and Medicine*, vol. 92, no. 1, pp. 21–28, 2019.

[421] F. Katsuki and C. Constantinidis, "Bottom-up and top-down attention: Different processes and overlapping neural systems", *The Neuroscientist*, vol. 20, no. 5, pp. 509–521, Oct. 1, 2014. DOI: 10.1177/1073858413514136.

[422] J. Theeuwes, P. Atchley, and A. F. Kramer, "On the time course of top-down and bottom-up control of visual attention", *Attention and Performance*, vol. 18, pp. 104–124, 2000. DOI: 10.7551/mitpress/1481.003.0011.

[423] T. Charitou, K. Bryan, and D. J. Lynn, "Using biological networks to integrate, visualize and analyze genomics data", *Genetics Selection Evolution*, vol. 48, no. 1, p. 27, Mar. 31, 2016. DOI: 10.1186/s12711-016-0205-1.

[424] A. Chiesi, "Network analysis", in *International Encyclopedia of the Social & Behavioral Sciences*, N. J. Smelser and P. B. Baltes, Eds., Oxford: Pergamon, Jan. 1, 2001, pp. 10 499–10 502. DOI: 10.1016/B0-08-043076-7/04211-X.

[425] J. Bailenson, R. Guadagno, E. Aharoni, A. Dimov, A. Beall, and J. Blascovich, "Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments", in *Proceedings of the 7th Annual International Workshop on PRESENCE*, 2004.

[426] R. E. Guadagno, J. Blascovich, J. N. Bailenson, and C. McCall, "Virtual humans and persuasion: The effects of agency and behavioral realism", *Media Psychology*, vol. 10, no. 1, pp. 1–22, 2007.

[427] S. Grover and R. Pea, "Computational thinking in k–12: A review of the state of the field", *Educational Researcher*, vol. 42, no. 1, pp. 38–43, Jan. 1, 2013. DOI: 10.3102/0013189X12463051.

[428] H. Gaspard, I. Häfner, C. Parrisius, U. Trautwein, and B. Nagengast, "Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject", *Contemporary Educational Psychology*, vol. 48, pp. 67–84, Jan. 1, 2017. DOI: 10.1016/j.cedpsych.2016.09.003.

[429] A. D. Schwanzer, U. Trautwein, O. Lüdtke, and H. Sydow, "Entwicklung eines instruments zur erfassung des selbstkonzepts junger erwachsener. [development of a questionnaire on young adults' self-concept.]", *Diagnostica*, vol. 51, no. 4, pp. 183–194, 2005. DOI: 10.1026/0012-1924.51.4.183.

[430] M. Román-González, J.-C. Pérez-González, and C. Jiménez-Fernández, "Which cognitive abilities underlie computational thinking? criterion validity of the computational thinking test", *Computers in Human Behavior*, vol. 72, pp. 678–691, Jul. 1, 2017. DOI: 10.1016/j.chb.2016.08.047.

[431] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. New York: Routledge, Jul. 1, 1988, 567 pp. DOI: 10.4324/9780203771587.

[432] R. C. Team, *R: A language and environment for statistical computing*, 2020.

[433] R. G. Fryer Jr. and S. D. Levitt, "An empirical analysis of the gender gap in mathematics", *American Economic Journal: Applied Economics*, vol. 2, no. 2, pp. 210–240, Apr. 2010. DOI: 10.1257/app.2.2.210.

[434] H. Plieninger and O. Dickhäuser, "The female fish is more responsive: Gender moderates the BFLPE in the domain of science", *Educational Psychology*, vol. 35, no. 2, pp. 213–227, Feb. 17, 2015. DOI: 10.1080/01443410.2013.814197.

[435] F. Preckel, M. Zeidner, T. Goetz, and E. J. Schleyer, "Female 'big fish' swimming against the tide: The 'big-fish-little-pond effect' and gender-ratio in special gifted classes", *Contemporary Educational Psychology*, vol. 33, no. 1, pp. 78–96, Jan. 1, 2008. DOI: 10.1016/j.cedpsych.2006.08.001.

[436] J. Tiedemann, "Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school", *Journal of Educational Psychology*, vol. 92, no. 1, pp. 144–151, 2000. DOI: 10.1037/0022-0663.92.1.144.

[437] D. Noton and L. Stark, "Eye movements and visual perception", *Scientific American*, vol. 224, no. 6, pp. 34–43, 1971.

[438] S. Cheryan, A. N. Meltzoff, and S. Kim, "Classrooms matter: The design of virtual classrooms influences gender disparities in computer science classes", *Computers & Education*, vol. 57, no. 2, pp. 1825–1835, Sep. 1, 2011. DOI: 10.1016/j.compedu.2011.02.004.

[439] J. N. Bailenson, K. Swinth, C. Hoyt, S. Persky, A. Dimov, and J. Blascovich, "The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments", *Presence: Teleoperators and Virtual Environments*, vol. 14, no. 4, pp. 379–393, Aug. 1, 2005. DOI: 10.1162/105474605774785235.

[440] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse, "The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '03, New York, NY, USA: ACM, Apr. 5, 2003, pp. 529–536. DOI: 10.1145/642611.642703.

[441] R. Ratan, D. Beyea, B. J. Li, and L. Graciano, "Avatar characteristics induce users' behavioral conformity with small-to-medium effect sizes: A meta-analysis of the proteus effect", *Media Psychology*, vol. 23, no. 5, pp. 651–675, Sep. 2, 2020. DOI: 10.1080/15213269.2019.1623698.

[442] R. E. Clark, "Reconsidering research on learning from media", *Review of Educational Research*, vol. 53, no. 4, pp. 445–459, Dec. 1, 1983. DOI: 10.3102/00346543053004445.

[443] Z. Merchant, E. T. Goetz, L. Cifuentes, W. Keeney-Kennicutt, and T. J. Davis, "Effectiveness of virtual reality-based instruction on students' learning outcomes in k-12 and higher education: A meta-analysis", *Computers & Education*, vol. 70, pp. 29–40, Jan. 1, 2014. DOI: 10.1016/j.compedu.2013.07.033.

[444] H. Jun, M. R. Miller, F. Herrera, B. Reeves, and J. N. Bailenson, "Stimulus sampling with 360-videos: Examining head movements, arousal, presence, simulator sickness, and preference on a large sample of participants and videos", *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1416–1425, Jul. 2022. DOI: 10.1109/TAFFC.2020.3004617.

[445] M. Roman Miller, H. Jun, and J. N. Bailenson, "Motion and meaning: Sample-level nonlinear analyses of virtual reality tracking data", in *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, Bari, Italy, Oct. 2021, pp. 147–152. DOI: 10.1109/ISMAR-Adjunct54149.2021.00039.